



8fw

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re U.S. Patent Application of

INOUE et al.

Application Number: 10/768,146

Filed: February 2, 2004

**For: DISK ARRAY DEVICE AND MAINTENANCE
METHOD FOR DISK ARRAY DEVICE**

Atty Docket No. WILL.0005

**Honorable Assistant Commissioner
for Patents
Washington, D.C. 20231**

LETTER

Sir:

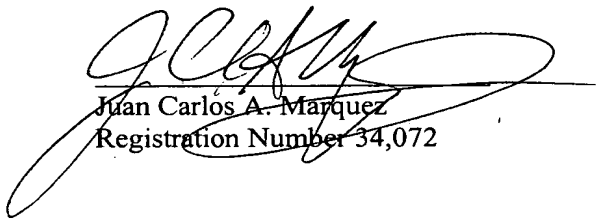
The below-identified communications are submitted in the above-captioned application or proceeding:

- | | | |
|---|---------------------------|---|
| <input checked="" type="checkbox"/> (X) | Priority Documents (1) | |
| <input checked="" type="checkbox"/> (X) | Request for Priority | <input type="checkbox"/> () Assignment Document |
| <input type="checkbox"/> () | Response to Missing Parts | <input type="checkbox"/> () Petition under 37 C.F.R. § 1.47(a) |
| | w/ signed Declaration | <input type="checkbox"/> () Check for |

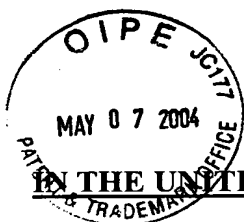
The Commissioner is hereby authorized to charge payment of any fees associated with this communication, including fees under 37 C.F.R. § 1.16 and 1.17 or credit any overpayment to **Deposit Account Number 08-1480**. A duplicate copy of this sheet is attached.

Respectfully submitted,

Stanley P. Fisher
Registration Number 24,344


Juan Carlos A. Marquez
Registration Number 34,072

REED SMITH LLP
3110 Fairview Park Drive
Suite 1400
Falls Church, Virginia 22042
(703) 641-4200
May 7, 2004



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re U.S. Patent Application of)
)
INOUE et al.)
)
Application Number: 10/768,146)
)
Filed: February 2, 2004)
)
For: DISK ARRAY DEVICE AND MAINTENANCE)
METHOD FOR DISK ARRAY DEVICE)
)
Atty Docket No. WILL.0005)

Honorable Assistant Commissioner
for Patents
Washington, D.C. 20231

**NOTICE OF PRIORITY
UNDER 35 U.S.C. 119
AND THE INTERNATIONAL CONVENTION**

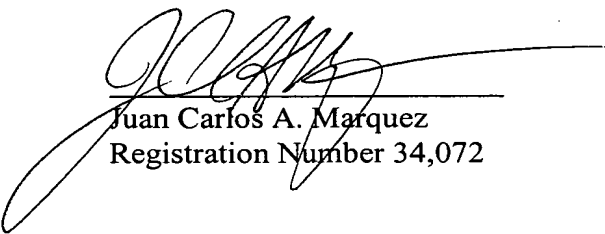
Sir:

In the matter of the above-captioned application for a United States patent, notice is hereby given that the Applicant claims the priority date of December 15, 2003, the filing date of the corresponding Japanese patent application 2003-416230.

The certified copy of corresponding Japanese patent application 2003-416230 is being submitted herewith. Acknowledgment of receipt of the certified copy is respectfully requested in due course.

Respectfully submitted,

Stanley P. Fisher
Registration Number 24,344



Juan Carlos A. Marquez
Registration Number 34,072

REED SMITH LLP
3110 Fairview Park Drive
Suite 1400
Falls Church, Virginia 22042
(703) 641-4200
May 7, 2004

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年12月15日
Date of Application:

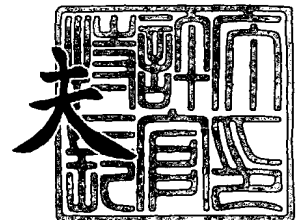
出願番号 特願2003-416230
Application Number:
[ST. 10/C]: [JP2003-416230]

出願人 株式会社日立製作所
Applicant(s):

2004年 1月27日

特許庁長官
Commissioner,
Japan Patent Office

今井 康夫



出証番号 出証特2004-3002964

【書類名】 特許願
【整理番号】 340301293
【あて先】 特許庁長官殿
【国際特許分類】 G06F 03/06
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 井上 雅生
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 奥元 勝博
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I
 D システム事業部内
 【氏名】 本間 久雄
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社日立製作所
【代理人】
 【識別番号】 100095371
 【弁理士】
 【氏名又は名称】 上村 輝之
【選任した代理人】
 【識別番号】 100089277
 【弁理士】
 【氏名又は名称】 宮川 長夫
【選任した代理人】
 【識別番号】 100104891
 【弁理士】
 【氏名又は名称】 中村 猛
【手数料の表示】
 【予納台帳番号】 043557
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1
 【包括委任状番号】 0110323

【書類名】 特許請求の範囲**【請求項 1】**

複数のクラスタが設けられたディスクアレイ装置において、
前記各クラスタは、
上位装置とのデータ授受を制御するチャンネルアダプタと、
記憶デバイスとのデータ授受を制御するディスクアダプタと、
キャッシュメモリを搭載するキャッシュメモリパッケージと、
装置構成及び装置動作に関する管理情報を記憶する基本制御メモリを搭載する基本制御メモリパッケージと、
前記キャッシュメモリの記憶構造に関するディレクトリ情報を記憶する拡張制御メモリを搭載する拡張制御メモリパッケージと、をそれぞれ備え、
前記管理情報は、前記各クラスタの前記各基本制御メモリにより多重化されて記憶されているディスクアレイ装置。

【請求項 2】

上位装置とのデータ授受を制御するチャンネルアダプタと、
記憶デバイスとのデータ授受を制御するディスクアダプタと、
前記チャンネルアダプタ及び前記ディスクアダプタにより使用され、データを記憶するキャッシュメモリを有するキャッシュメモリパッケージと、
前記チャンネルアダプタ及び前記ディスクアダプタにより使用され、制御情報を記憶する制御メモリを有する複数の制御メモリパッケージと、を備え、
前記制御情報は、第 1 の制御情報と第 2 の制御情報とを含み、
前記第 1 の制御情報は、互いに異なる制御メモリにそれぞれ記憶されて多重化されており、
前記第 2 の制御情報は、前記第 1 の制御情報をそれぞれ記憶する前記各制御メモリとは異なる別の制御メモリに記憶されているディスクアレイ装置。

【請求項 3】

前記第 1 の制御情報は、ディスクアレイ装置の動作を制御するために用いられる管理情報であり、
前記第 2 の制御情報は、前記キャッシュメモリの記憶構造に関する記憶構造情報である請求項 2 に記載のディスクアレイ装置。

【請求項 4】

前記複数の制御メモリのいずれかに障害が発生した場合に、障害の発生した制御メモリに記憶されている情報を復旧させる保守制御部を設け、
前記保守制御部は、前記第 1 の制御情報を記憶する制御メモリの記憶領域を利用して、前記障害の発生した制御メモリに記憶されている情報を復旧させるものである請求項 2 に記載のディスクアレイ装置。

【請求項 5】

前記保守制御部は、前記第 2 の制御情報を記憶する制御メモリに障害が発生した場合に、前記第 1 の制御情報を記憶する制御メモリの空き領域に再構築可能な分だけ前記第 2 の制御情報を復旧させる請求項 4 に記載のディスクアレイ装置。

【請求項 6】

前記保守制御部は、前記障害の発生した制御メモリの記憶構造が前記障害の発生前後で異なるように、前記障害の発生した制御メモリに記憶されている情報を復旧させる請求項 5 に記載のディスクアレイ装置。

【請求項 7】

前記保守制御部は、(1) 前記第 1 の制御情報を記憶する制御メモリに利用可能な空き領域が存在しない場合に実行可能な第 1 の保守モードと、(2) 前記第 1 の制御情報を記憶する制御メモリに利用可能な空き領域が所定値以上存在する場合に実行可能な第 2 の保守モードと、(3) 前記第 1 の制御情報を記憶する制御メモリに利用可能な空き領域が前記所定値未満だけ存在する場合に実行可能な第 3 の保守モードとのうち、少なくともいずれ

れか1つの保守モードを実行するものであり、

(1) 前記第1の保守モードは、

(1-1) 前記第1の制御情報を記憶する制御メモリに障害が発生した場合は、この障害の発生した制御メモリが正常品と交換された場合に、他の制御メモリに多重化されている前記第1の制御情報を前記交換された制御メモリにコピーさせることにより前記第1の制御情報を復旧し、

(1-2) 前記第2の制御情報を記憶する制御メモリに障害が発生した場合は、前記第1の制御情報を記憶する制御メモリに上書きで前記第2の制御情報を再構築し、

前記障害の発生した制御メモリが正常品と交換された場合に、この交換された制御メモリに前記第2の制御情報を再構築し、

前記第2の制御情報が上書きされた制御メモリに、他の制御メモリにより多重化されている前記第1の制御情報をコピーさせることにより前記第1の制御情報を復旧するものであり、

(2) 前記第2の保守モードは、

(2-1) 前記第1の制御情報を記憶する制御メモリに障害が発生した場合は、この障害の発生した制御メモリが正常品と交換された場合に、他の制御メモリに多重化されている前記第1の制御情報を前記交換された制御メモリにコピーさせることにより前記第1の制御情報を復旧し、

(2-2) 前記第2の制御情報を記憶する制御メモリに障害が発生した場合は、前記第1の制御情報を記憶する制御メモリの空き領域に前記第2の制御情報を再構築し、この障害の発生した制御メモリの正常品との交換を許可するものであり、

(3) 前記第3の保守モードは、

(3-1) 前記第1の制御情報を記憶する制御メモリに障害が発生した場合は、この障害の発生した制御メモリが正常品と交換された場合に、他の制御メモリに多重化されている前記第1の制御情報を前記交換された制御メモリにコピーさせることにより前記第1の制御情報を復旧し、

(3-2) 前記第2の制御情報を記憶する制御メモリに障害が発生した場合は、前記第1の制御情報を記憶する制御メモリの空き領域に再構築可能な範囲だけ前記第2の制御情報を部分的に再構築し、

前記障害の発生した制御メモリが正常品と交換された場合に、この交換された制御メモリに前記第2の制御情報のうち再構築されていない残余の部分を再構築するものである、請求項4に記載のディスクアレイ装置。

【請求項8】

前記保守制御部は、情報復旧作業に係わる制御メモリの記憶構造をそのまま保持するものであり、かつ、前記第2の保守モードを実行した場合において、

(2-1A) 前記第1の制御情報を記憶する制御メモリに障害が発生した場合は、前記交換された制御メモリに前記第2の制御情報を再構築し、

前記障害の発生した制御メモリが正常品と交換された場合に、この交換された制御メモリに他の制御メモリに多重化されている前記第1の制御情報をコピーさせることにより前記第1の制御情報を復旧し、

(2-2A) 前記交換された制御メモリに障害が発生した場合は、正常品との交換を許可する、ようになっている請求項7に記載のディスクアレイ装置。

【請求項9】

前記保守制御部は、情報復旧作業に係わる制御メモリの記憶構造をそのまま保持するものであり、かつ、前記第3の保守モードを実行した場合において、

(3-1A) 前記第1の制御情報を記憶する制御メモリに障害が発生した場合は、前記交換された制御メモリに前記第2の制御情報を再構築し、

前記障害の発生した制御メモリが正常品と交換された場合に、この交換された制御メモリに他の制御メモリに多重化されている前記第1の制御情報をコピーさせることにより前記第1の制御情報を復旧し、

(3-2A) 前記交換された制御メモリに障害が発生した場合は、正常品との交換を許可し、正常品と交換された場合に、この交換された制御メモリに、前記第2の制御情報のうち前記残余の部分を再構築させるようになっている、請求項7に記載のディスクアレイ装置。

【請求項10】

前記保守制御部は、前記第3の保守モードを実行する場合において、前記キャッシュメモリの使用領域を、前記第1の制御情報を記憶する制御メモリの空き領域に部分的に再構築された前記第2の制御情報で管理可能な範囲に制限させる請求項7または請求項9のいずれかに記載のディスクアレイ装置。

【請求項11】

上位装置とのデータ授受を制御するチャネルアダプタと、記憶デバイスとのデータ授受を制御するディスクアダプタと、前記チャネルアダプタ及び前記ディスクアダプタにより使用され、データをそれぞれ記憶する複数のキャッシュメモリと、前記チャネルアダプタ及び前記ディスクアダプタにより使用され、装置構成及び装置動作に関する管理情報をそれぞれ記憶する複数の第1制御メモリと、前記チャネルアダプタ及び前記ディスクアダプタにより使用され、前記各キャッシュメモリの記憶構造に関する記憶構造情報をそれぞれ記憶する複数の第2制御メモリとを備えたディスクアレイ装置の保守方法であって、

前記管理情報は、前記各第1制御メモリにそれぞれ記憶させて多重化しておき、

前記各第1制御メモリ及び前記各第2制御メモリのいずれかに障害が発生したか否かを検出する障害検出ステップと、

前記障害が検出された場合は、前記第1制御メモリの記憶領域を利用して、前記障害の発生した制御メモリに記憶されている情報を復旧させる保守ステップとを、含んだディスクアレイ装置の保守方法。

【請求項12】

前記保守ステップは、(1) 前記第1制御メモリに利用可能な空き領域が存在しない場合に実行可能な第1の保守モードと、(2) 前記第1制御メモリに利用可能な空き領域が所定値以上存在する場合に実行可能な第2の保守モードと、(3) 前記第1制御メモリに利用可能な空き領域が前記所定値未満だけ存在する場合に実行可能な第3の保守モードとのうち、いずれか1つの保守モードを排他的に実行するものであり、

(1) 前記第1の保守モードは、

(1-1) 前記第1制御メモリに障害が発生した場合は、この障害の発生した第1制御メモリが正常品と交換された場合に、他の第1制御メモリに多重化されている前記管理情報を前記交換された第1制御メモリにコピーさせるステップと、

(1-2) 前記第2制御メモリに障害が発生した場合は、前記第1制御メモリに上書きで前記記憶構造情報を再構築し、

前記障害の発生した第2制御メモリが正常品と交換された場合に、この交換された第2制御メモリに前記記憶構造情報を再構築し、

前記記憶構造情報が上書きされた第1制御メモリに、他の第1制御メモリにより多重化されている前記管理情報をコピーさせるステップと、を含んでおり、

(2) 前記第2の保守モードは、

(2-1) 前記第1制御メモリに障害が発生した場合は、この障害の発生した第1制御メモリが正常品と交換された場合に、他の第1制御メモリに多重化されている前記管理情報を前記交換された制御メモリにコピーさせるステップと、

(2-2) 前記第2制御メモリに障害が発生した場合は、前記第1制御メモリの空き領域に前記記憶構造情報を再構築するステップ、及びこの障害の発生した第2制御メモリの正常品との交換を許可するステップと、を含んでおり、

(3) 前記第3の保守モードは、

(3-1) 前記第1制御メモリに障害が発生した場合は、この障害の発生した第1制御メモリが正常品と交換された場合に、他の第1制御メモリに多重化されている前記管理情報を前記交換された第1制御メモリにコピーさせるステップと、

(3-2) 前記第2制御メモリに障害が発生した場合は、前記第1制御メモリの空き領域に再構築可能な範囲だけ前記記憶構造情報を部分的に再構築するステップと、

前記障害の発生した第2制御メモリが正常品と交換された場合に、この交換された第2制御メモリに前記記憶構造情報のうち再構築されていない残余の部分を再構築するステップと、を含んでいる請求項11に記載のディスクアレイ装置の保守方法。

【請求項13】

前記保守ステップは、前記第2の保守モードを実行した場合において、

(2-1A) 前記第1制御メモリに障害が発生した場合は、前記交換された第2制御メモリに前記記憶構造情報を再構築するステップと、

前記障害の発生した第1制御メモリが正常品と交換された場合に、この交換された第1制御メモリに他の第1制御メモリに多重化されている前記管理情報をコピーさせるステップと、

(2-2A) 前記交換された第2制御メモリに障害が発生した場合は、正常品との交換を許可するステップとを、さらに含んでいる請求項12に記載のディスクアレイ装置の保守方法。

【請求項14】

前記保守ステップは、前記第3の保守モードを実行した場合において、

(3-1A) 前記第1制御メモリに障害が発生した場合は、前記交換された第2制御メモリに前記記憶構造情報を再構築し、

前記障害の発生した第1制御メモリが正常品と交換された場合に、この交換された第1制御メモリに他の第1制御メモリに多重化されている前記管理情報をコピーさせるステップと、

(3-2A) 前記交換された第2制御メモリに障害が発生した場合は、正常品との交換を許可するステップと、

正常品と交換された場合に、この交換された第2制御メモリに、前記記憶構造情報のうち前記残余の部分を再構築させるステップとを、さらに含んでいる請求項12に記載のディスクアレイ装置の保守方法。

【書類名】 明細書**【発明の名称】** ディスクアレイ装置及びディスクアレイ装置の保守方法**【技術分野】****【0001】**

本発明は、ディスクアレイ装置及びディスクアレイ装置の保守方法に関する。

【背景技術】**【0002】**

ディスクアレイ装置は、例えば、多数のディスクドライブをアレイ状に配設し、RAID (Redundant Array of Independent Inexpensive Disks) に基づいて構築されている。各ディスク装置が有する物理的な記憶領域上には、論理的な記憶領域である論理ボリュームが形成されている。ホストコンピュータは、ディスクアレイ装置に対して所定形式の書き込みコマンド又は読出しコマンドを発行することにより、所望のデータの読み書きを行うことができる。

【0003】

ディスクアレイ装置には、ディスクドライブに記憶したデータの消失等を防止するために、種々の防御策が施されている。1つは、RAID構成の採用である。例えば、RAID 1～6等として知られている冗長記憶構造をディスクアレイ装置が採用することにより、データ消失の可能性が低減される。

【0004】

さらに、ディスクアレイ装置では、物理的構成の二重化も行われている。例えば、ディスクアレイ装置では、ホストコンピュータとの間のデータ通信を行う上位インターフェース回路や各ディスクドライブとの間のデータ通信を行う下位インターフェース回路等の主要部を複数設けて多重化している。また、これら各主要部間をそれぞれ接続する経路や、各主要部に電力を供給する電源等も複数設けられている。

【0005】

また、従来のディスクアレイ装置は、共有メモリとキャッシュメモリとを備えており、共有メモリには制御情報を記憶し、キャッシュメモリにはデータを記憶するようになっている(特許文献1)。従来技術では、1つのメモリパッケージ内に共有メモリとキャッシュメモリの両方を搭載している。また、キャッシュ容量を補うために、セクタパッケージ内にもキャッシュメモリを搭載している。

【特許文献1】 特開 2000-339101号公報**【発明の開示】****【発明が解決しようとする課題】****【0006】**

上位装置から書き込みを要求され、ディスクドライブにまだ書き込まれていないユーザデータは、複数のキャッシュメモリに記憶させて二重化する必要がある。データ保証を行うためである。また、ディスクアレイ装置の動作等を制御するための制御情報も、冗長性確保の観点から、二重化される必要がある。制御情報を二重化することにより、一方の制御情報が障害発生等によって利用不能となった場合でも、他方の制御情報に基づいて、上位装置からの読み書き要求を処理することができる。

【0007】

しかし、制御情報を記憶する制御メモリとユーザデータを記憶するキャッシュメモリとが同一パッケージに設けられている構成では、制御メモリまたはキャッシュメモリのいずれかに障害が発生した場合に、障害の発生したパッケージ全体の保守交換作業が完了するまでの期間中、応答性が低下する。

【0008】

いずれのパッケージにも障害が発生していない正常状態で上位装置からの書き込み要求があった場合は、データをキャッシュメモリに記憶した時点で、書き込み完了を上位装置に報告することができる。書き込みを要求されたデータは、キャッシュメモリに二重化されてデータ保証が行われているため、ディスクドライブにデータを書き込む前に上位装置に書込

み完了を報告しても、何ら不都合はない。

【0009】

これに対し、いずれか一方のパッケージ全体が閉塞処理された場合は、正常に稼働している他方のパッケージに搭載された制御メモリ及びキャッシュメモリのみしか利用することができない。従って、この場合は、二重化によるデータ保証を行うことができないため、書込みを要求されたデータを単一のキャッシュメモリに記憶させた時点では、上位装置に対して書込み完了を報告することはできない。そこで、いずれか一方のパッケージが使用不能な場合は、書込みを要求されたデータをディスクドライブに書き込んだ後で、上位装置に書込み完了を報告するようになっている。この応答動作は、例えば、「疑似スルー動作」と呼ばれる。

【0010】

疑似スルー動作時でも、ディスクアレイ装置は正常に稼働を続けることができる。しかし、上位装置への応答が遅れるため、ディスクアレイ装置の性能が低下する。疑似スルー動作は、障害の発生したパッケージが交換されて、必要な作業が完了するまでの間ずっと継続する。従って、保守回復に要する全期間にわたって疑似スルー動作が行われ、ディスクアレイ装置の応答性が低下する。

【0011】

本発明の1つの目的は、制御メモリまたはキャッシュメモリのいずれかに障害が発生した場合でも、他のメモリに影響を与えることなく、保守回復作業を行うことができるディスクアレイ装置及びディスクアレイ装置の保守方法を提供することにある。本発明の1つの目的は、障害が発生した場合でも応答性の低下を抑制できるようにしたディスクアレイ装置及びディスクアレイ装置の保守方法を提供することにある。本発明の1つの目的は、制御情報の種類に応じて制御情報を複数の制御メモリにそれぞれ格納させることにより、制御メモリを有効に利用して保守回復作業を行うことができるようにしたディスクアレイ装置及びディスクアレイ装置の保守方法を提供することにある。本発明の他の目的は、後述する実施の形態の記載から明らかになるであろう。

【課題を解決するための手段】

【0012】

上記課題を解決すべく、本発明に従うディスクアレイ装置は、上位装置とのデータ授受を制御するチャンネルアダプタと、記憶デバイスとのデータ授受を制御するディスクアダプタと、チャンネルアダプタ及びディスクアダプタにより使用され、データを記憶するキャッシュメモリを有するキャッシュメモリパッケージと、チャンネルアダプタ及びディスクアダプタにより使用され、制御情報を記憶する制御メモリを有する複数の制御メモリパッケージと、を備えている。制御情報は、第1の制御情報と第2の制御情報とを含んでいる。そして、第1の制御情報は、互いに異なる制御メモリにそれぞれ記憶されて多重化されており、第2の制御情報は、第1の制御情報をそれぞれ記憶する各制御メモリとは異なる別の制御メモリに記憶されている。

【0013】

制御メモリパッケージとキャッシュメモリパッケージとは別々のパッケージとして構成されるため、いずれかのパッケージに障害が発生した場合でも、他のパッケージに影響を与えずに独立して、保守回復作業を行うことができる。また、第1の制御情報は多重化して管理されるため、ディスクアレイ装置の動作を制御するために用いられる管理情報を第1の制御情報として用いることができる。また、第2の制御情報としては、キャッシュメモリの記憶構造に関する記憶構造情報を用いることができる。

【0014】

本発明の一態様では、複数の制御メモリのいずれかに障害が発生した場合に、障害の発生した制御メモリに記憶されている情報を復旧させる保守制御部を設けている。保守制御部は、第1の制御情報を記憶する制御メモリの記憶領域を利用して、障害の発生した制御メモリに記憶されている情報を復旧させる。

【0015】

第1の制御情報としての管理情報は多重管理が要請され、第2の制御情報としての記憶構造情報は多重管理が要請されない、という技術的性質をそれぞれ有する。もしも障害等によって記憶構造情報が使用不能となった場合は、疑似スルー動作を行いながら、上位装置からの書込み要求を処理する。しかし、記憶構造情報が回復した場合は、正常状態に復帰し、疑似スルー動作は停止する。ここで、保守制御部は、第1の制御情報を記憶する制御メモリの記憶領域を利用して第2の制御情報を回復させることができる。従って、第2の制御情報を記憶する制御メモリが正常品に交換される前に、第2の制御情報を回復させて、疑似する一動作を廃した正常な動作に戻すことができる。

【0016】

本発明の一態様では、保守制御部は、第2の制御情報を記憶する制御メモリに障害が発生した場合に、第1の制御情報を記憶する制御メモリの空き領域に再構築可能な分だけ第2の制御情報を復旧させる。

【0017】

また、本発明の一態様では、保守制御部は、障害の発生した制御メモリの記憶構造が障害の発生前後で異なるように、障害の発生した制御メモリに記憶されている情報を復旧させる。即ち、保守制御部は、障害発生前の初期状態と同一の状態を形成することによって保守回復を行うのではなく、データ保証が可能な状態まで回復させることができる。従って、保守回復に要する期間を短縮することができる。

【0018】

本発明の一態様では、保守制御部は、(1)第1の制御情報を記憶する制御メモリに利用可能な空き領域が存在しない場合に実行可能な第1の保守モードと、(2)第1の制御情報を記憶する制御メモリに利用可能な空き領域が所定値以上存在する場合に実行可能な第2の保守モードと、(3)第1の制御情報を記憶する制御メモリに利用可能な空き領域が所定値未満だけ存在する場合に実行可能な第3の保守モードとのうち、少なくともいずれか1つの保守モードを実行するようになっている。(1)第1の保守モードは、(1-1)第1の制御情報を記憶する制御メモリに障害が発生した場合は、この障害の発生した制御メモリが正常品と交換された場合に、他の制御メモリに多重化されている第1の制御情報を、交換された制御メモリにコピーさせることにより第1の制御情報を復旧し、(1-2)第2の制御情報を記憶する制御メモリに障害が発生した場合は、第1の制御情報を記憶する制御メモリに上書きで第2の制御情報を再構築し、障害の発生した制御メモリが正常品と交換された場合に、この交換された制御メモリに第2の制御情報を再構築し、第2の制御情報が上書きされた制御メモリに、他の制御メモリにより多重化されている第1の制御情報をコピーさせることにより第1の制御情報を復旧するモードである。(2)第2の保守モードは、(2-1)第1の制御情報を記憶する制御メモリに障害が発生した場合は、この障害の発生した制御メモリが正常品と交換された場合に、他の制御メモリに多重化されている第1の制御情報を交換された制御メモリにコピーさせることにより第1の制御情報を復旧し、(2-2)第2の制御情報を記憶する制御メモリに障害が発生した場合は、第1の制御情報を記憶する制御メモリの空き領域に第2の制御情報を再構築し、この障害の発生した制御メモリの正常品との交換を許可するモードである。(3)第3の保守モードは、(3-1)第1の制御情報を記憶する制御メモリに障害が発生した場合は、この障害の発生した制御メモリが正常品と交換された場合に、他の制御メモリに多重化されている第1の制御情報を交換された制御メモリにコピーさせることにより第1の制御情報を復旧し、(3-2)第2の制御情報を記憶する制御メモリに障害が発生した場合は、第1の制御情報を記憶する制御メモリの空き領域に再構築可能な範囲だけ第2の制御情報を部分的に再構築し、障害の発生した制御メモリが正常品と交換された場合に、この交換された制御メモリに第2の制御情報のうち再構築されていない残余の部分再構築するモードである。

【0019】

また、本発明の一態様では、一度復旧作業を行った後で、さらに別の障害が発生した場合をも考慮している。即ち、保守制御部は、情報復旧作業に係わる制御メモリの記憶構造

をそのまま保持するものであり、かつ、第2の保守モードを実行した場合において、(2-1A) 第1の制御情報を記憶する制御メモリに障害が発生した場合は、交換された制御メモリに第2の制御情報を再構築し、障害の発生した制御メモリが正常品と交換された場合に、この交換された制御メモリに他の制御メモリに多重化されている第1の制御情報をコピーさせることにより第1の制御情報を復旧し、(2-2A) 交換された制御メモリに障害が発生した場合は、正常品との交換を許可する。

【0020】

あるいは、保守制御部は、情報復旧作業に係わる制御メモリの記憶構造をそのまま保持するものであり、かつ、第3の保守モードを実行した場合において、(3-1A) 第1の制御情報を記憶する制御メモリに障害が発生した場合は、交換された制御メモリに第2の制御情報を再構築し、障害の発生した制御メモリが正常品と交換された場合に、この交換された制御メモリに他の制御メモリに多重化されている第1の制御情報をコピーさせることにより第1の制御情報を復旧し、(3-2A) 交換された制御メモリに障害が発生した場合は、正常品との交換を許可し、正常品と交換された場合に、この交換された制御メモリに、第2の制御情報のうち前記残余の部分を再構築させる。

【0021】

さらに、本発明の一態様では、保守制御部は、第3の保守モードを実行する場合において、キャッシュメモリの使用領域を、第1の制御情報を記憶する制御メモリの空き領域に部分的に再構築された第2の制御情報で管理可能な範囲に制限させる。

【発明を実施するための最良の形態】

【0022】

以下、図1～図21に基づき、本発明の実施の形態を説明する。本実施形態では、複数のクラスタが設けられたディスクアレイ装置が開示されており、各クラスタは、上位装置とのデータ授受を制御するチャネルアダプタと、記憶デバイスとのデータ授受を制御するディスクアダプタと、キャッシュメモリを搭載するキャッシュメモリパッケージと、装置構成及び装置動作に関する管理情報を記憶する基本制御メモリを搭載する基本制御メモリパッケージと、キャッシュメモリの記憶構造に関するディレクトリ情報を記憶する拡張制御メモリを搭載する拡張制御メモリパッケージと、をそれぞれ備えている。そして、管理情報は、各クラスタの各基本制御メモリにより多重化されて記憶されている。

【0023】

また、本実施形態では、ディスクアレイ装置の保守方法が開示されている。即ち、上位装置とのデータ授受を制御するチャネルアダプタと、記憶デバイスとのデータ授受を制御するディスクアダプタと、チャネルアダプタ及びディスクアダプタにより使用され、データをそれぞれ記憶する複数のキャッシュメモリと、チャネルアダプタ及びディスクアダプタにより使用され、装置構成及び装置動作に関する管理情報をそれぞれ記憶する複数の第1制御メモリと、チャネルアダプタ及びディスクアダプタにより使用され、各キャッシュメモリの記憶構造に関する記憶構造情報をそれぞれ記憶する複数の第2制御メモリとを備えたディスクアレイ装置の保守方法であって、管理情報は、各第1制御メモリにそれぞれ記憶させて多重化しておき、各第1制御メモリ及び各第2制御メモリのいずれかに障害が発生したか否かを検出する障害検出ステップと、障害が検出された場合は、第1制御メモリの記憶領域を利用して、障害の発生した制御メモリに記憶されている情報を復旧させる保守ステップとを、含んでいる。

【実施例1】

【0024】

図1は、ディスクアレイ装置10の概略構成を示すブロック図である。ディスクアレイ装置10は、通信ネットワークCN1を介して、複数のホストコンピュータ1と双方向通信可能に接続されている。ここで、通信ネットワークCN1は、例えば、LAN (Local Area Network)、SAN (Storage Area Network)、インターネット等である。LANを用いる場合、ホストコンピュータ1とディスクアレイ装置10との間のデータ転送は、TCP/IP (Transmission Control Protocol/Internet Protocol) プロトコルに従って行われ

る。SANを用いる場合、ホストコンピュータ1とディスクアレイ装置10とは、ファイバチャネルプロトコルに従ってデータ転送を行う。また、ホストコンピュータ1がメインフレームの場合は、例えば、FICON (Fibre Connection:登録商標)、ESCON (Enterprise System Connection:登録商標)、ACONARC (Advanced Connection Architecture:登録商標)、FIBARC (Fibre Connection Architecture:登録商標)等の通信プロトコルに従ってデータ転送が行われる。

【0025】

各ホストコンピュータ1は、例えば、サーバ、パーソナルコンピュータ、ワークステーション、メインフレーム等として実現されるものである。例えば、各ホストコンピュータ1は、図外に位置する複数のクライアント端末と別の通信ネットワークを介して接続されている。各ホストコンピュータ1は、例えば、各クライアント端末からの要求に応じて、ディスクアレイ装置10にデータの読み書きを行うことにより、各クライアント端末へのサービスを提供する。

【0026】

ディスクアレイ装置10は、それぞれ後述するように、複数のチャネルアダプタ(以下、CHAと略記)20A~20Dと、複数のディスクアダプタ(以下、DKAと略記)30A~30Dと、複数のディスクドライブ41と、複数のキャッシュパッケージ50A~50D、60A~60Dと、接続部71、72とを備えている。また、ディスクアレイ装置10内には、2つのクラスタ11A、11Bがそれぞれ形成されている。各クラスタ11A、11Bは、基本的に同一の物理的構成を有するが、それぞれが管理するデータ等は相違する場合がある。

【0027】

CHA20A~20D及びDKA30A~30Dは、例えば、プロセッサやメモリ等が実装されたプリント基板と、メモリに格納された制御プログラムとをそれぞれ備えており、これらのハードウェアとソフトウェアとの協働作業によって所定の機能を実現するようになっている。

【0028】

以下の説明では、個々の要素を特に区別しない場合は、「CHA20」、「DKA30」、「キャッシュパッケージ50」、「キャッシュパッケージ60」、「キャッシュメモリ51」、「共有メモリ61」等のように、符号に添えられたアルファベットを省略して説明する。

【0029】

ディスクアレイ装置10には、例えば、4個や8個等のように、複数のCHA20が設けられている。各CHA20は、例えば、オープン系用CHA、メインフレーム系用CHA等のように、ホストコンピュータ1の種類に応じて用意される。各CHA20は、ホストコンピュータ1との間のデータ転送を制御する。各CHA20は、それぞれプロセッサ部、データ通信部及びローカルメモリ部を備えている(いずれも不図示)。

【0030】

各CHA20は、それぞれに接続されたホストコンピュータ1から、データの読み書きを要求するコマンド及びデータを受信し、ホストコンピュータ1から受信したコマンドに従って動作する。DKA30の動作も含めて先に説明すると、例えば、CHA20は、ホストコンピュータ1からデータの読出し要求を受信すると、読出しコマンドを共有メモリ61に記憶させる。DKA30は、共有メモリ61を随時参照しており、未処理の読出しコマンドを発見すると、ディスクドライブ41からデータを読み出して、キャッシュメモリ51に記憶させる。CHA20は、キャッシュメモリ51に移されたデータを読み出し、コマンド発行元のホストコンピュータ1に送信する。

【0031】

また例えば、CHA20は、ホストコンピュータ1からデータの書き込み要求を受信すると、書き込みコマンドを共有メモリ61に記憶させると共に、受信したデータ(ユーザデータ)をキャッシュメモリ51に記憶させる。ここで、ホストコンピュータ1から書き込みを

要求されたデータは、ディスクドライブ 41 に書き込まれていない「ダーティデータ」であるため、複数のキャッシュメモリ 51 にそれぞれ記憶され、多重化される。CHA20 は、キャッシュメモリ 51 にデータを記憶した後、ホストコンピュータ 1 に対して書き込み完了を報告する。そして、DKA30 は、共有メモリ 61 に記憶された書き込みコマンドに従って、キャッシュメモリ 51 に記憶されたデータを読み出し、所定のディスクドライブ 41 に記憶させる。ディスクドライブ 41 に書き込まれたデータは、「ダーティデータ」から「グリーンデータ」に属性が変化し、キャッシュメモリ 51 による多重管理から解放される。

【0032】

ここで、本明細書において、「ダーティデータ」とは、ディスクドライブ 41 に書き込まれていないデータであって、複数のキャッシュメモリ 51 による多重管理が要請されるデータを意味する。また、本実施例において「クリーンデータ」とは、ディスクドライブ 41 に書き込まれているデータであって、複数のキャッシュメモリ 51 による多重管理が要請されないデータを意味する。

【0033】

各DKA30は、ディスクアレイ装置10内に例えば4個や8個等のように複数個設けられている。各DKA30は、各ディスクドライブ41との間のデータ通信を制御するもので、それぞれプロセッサ部と、データ通信部と、ローカルメモリ等を備えている（いずれも不図示）。各DKA30と各ディスクドライブ41とは、例えば、SAN等の通信ネットワークを介して接続されており、ファイバチャネルプロトコルに従ってブロック単位のデータ転送を行う。各DKA30は、ディスクドライブ41の状態を随時監視しており、この監視結果は内部ネットワークを介してSVP2に送信される。

【0034】

ディスクアレイ装置10は、記憶装置40を備えている。記憶装置40は、多数のディスクドライブ41から構成される。各ディスクドライブ41は、例えば、ハードディスクドライブ（HDD）や半導体メモリ装置等として実現可能である。記憶装置40は、ディスクアレイ装置10の内部に設けられている必要はなく、外部に存在してもよい。また、ディスクアレイ装置10に固有の記憶装置40である必要はなく、旧型の記憶装置や他社の記憶装置であってもよい。記憶装置40が他社製品等の場合は、ディスクアレイ装置10の論理デバイス（LDEV）または論理デバイスの下に設けられる仮想デバイス（VDEV）に他社製の記憶装置40をマッピングし、あたかもディスクアレイ装置10自身の記憶装置であるかのようにして用いることができる。

【0035】

ここで、例えば、4個のディスクドライブ41によって1つのRAIDグループを構成することができる。RAIDグループとは、例えばRAID5（RAID5に限定されない）に従って、データの冗長記憶を実現するディスクグループである。各RAIDグループにより提供される物理的な記憶領域の上には、論理的な記憶領域である論理ボリューム42（LUまたはLDEV）を少なくとも1つ以上設定可能である。

【0036】

メモリ部12は、複数のキャッシュパッケージ50、60から構成されている。一方のキャッシュパッケージ50は、キャッシュメモリ51を搭載し、他方のキャッシュパッケージ60は、共有メモリ61を搭載している。従って、他方のキャッシュパッケージ60を共有メモリパッケージまたは制御メモリパッケージと呼ぶことも可能である。ここで、キャッシュメモリ51には、データが記憶される。共有メモリ61には、制御情報が記憶される。共有メモリ61は、不揮発メモリであることが好ましい。また、キャッシュメモリ51も不揮発メモリとして構成することができる。後述のように、本実施例では、制御情報の技術的性質に着目し、制御情報の属性に基づいて情報を分類管理している。

【0037】

また、キャッシュパッケージ50、60は、基本メモリ部（図中「Basic」）と増設メモリ部（図中「Option」）とに分類される。一例として、基本メモリ部はディスクアレイ

装置 10 に標準で搭載されているものであり、増設メモリ部は必要に応じて増設されるオプション部品である。基本メモリ部及び増設メモリ部は、それぞれキャッシュメモリ 51 を搭載したキャッシュパッケージ 50 と共有メモリ 61 を搭載したキャッシュパッケージ 60 とを備えている。

【0038】

各クラスタ 11A, 11B 毎に、それぞれ基本メモリ部と増設メモリ部とが設けられている。クラスタ 11A の基本メモリ部は、キャッシュパッケージ 50A 及びキャッシュパッケージ 60A から構成され、クラスタ 11A の増設メモリ部は、キャッシュパッケージ 50B 及びキャッシュパッケージ 60B から構成される。クラスタ 11B の基本メモリ部は、キャッシュパッケージ 50C 及びキャッシュパッケージ 60C から構成され、クラスタ 11B の増設メモリ部は、キャッシュパッケージ 50D 及びキャッシュパッケージ 60D から構成される。従って、各クラスタ 11A, 11B は、それぞれ 2 つずつのキャッシュメモリ 51 及び共有メモリ 61 を備えている。

【0039】

接続部 71 は、各 CHA 20 と各キャッシュパッケージ 50, 60 とをそれぞれ接続するものである。接続部 72 は、各 DKA 30 と各キャッシュパッケージ 50, 60 とをそれぞれ接続するものである。これにより、全ての CHA 20, DKA 30 は、全てのキャッシュパッケージ 50, 60 にそれぞれアクセス可能である。

【0040】

SVP (Service Processor) 2 は、ディスクアレイ装置 10 の管理及び監視を行うためのコンピュータ装置である。SVP 2 は、ディスクアレイ装置 10 内に設けられた LAN 等の内部ネットワークを介して、各 CHA 20 及び各 DKA 30 等から各種の環境情報や性能情報等を収集する。SVP 2 が収集する情報としては、例えば、装置構成、電源アラーム、温度アラーム、入出力速度 (IOPS) 等が挙げられる。システム管理者は、SVP 2 の提供するユーザインターフェースを介して、RAID 構成の設定、各種パッケージ (CHA, DKA、ディスクドライブ等) の閉塞処理等を行うことができる。

【0041】

図 2 は、キャッシュパッケージ 60 の概略構成を示す。制御用のキャッシュパッケージ 60 は、例えば、共有メモリ 61 と、メモリコントロール回路 (図中「CTRL」) 62 と、アダプタ (図中「ADP」) 63 とを備えている。共有メモリ 61 は、例えば、メモリスロットに着脱可能に装着された少なくとも 1 つ以上の不揮発メモリモジュールから構成される。メモリコントロール回路 62 は、例えば、共有メモリ 61 へのデータ書込み及び読出しを制御する。アダプタ 63 は、例えば、接続部 71, 72 を介して各 CHA 20, 各 DKA 30 にそれぞれ接続される。なお、図示は省略するが、キャッシュパッケージ 50 も、前記同様に、キャッシュメモリ 51 と、メモリコントロール回路と、アダプタとを備えている。

【0042】

上述の通り、本実施例では、制御情報を記憶するための共有メモリ 61 を合計 4 個備えている。ここで、制御情報は、例えば、管理情報とディレクトリ情報とに分類することができる。本明細書における管理情報とは、例えば、ディスクアレイ装置 10 の装置構成や装置動作等を制御するために必要な制御情報と定義することが可能である。従って、管理情報は、多重化して管理する必要がある。各クラスタ 11A, 11B で管理情報がそれぞれ管理されているが、管理情報は装置全体の動作等を制御するものであり、クラスタには依存しない。別々のクラスタにそれぞれ存在する管理情報の内容は、一致するように管理されている。

【0043】

一方、本明細書におけるディレクトリ情報とは、例えば、キャッシュメモリ 51 の記憶構造等を示す制御情報と、あるいは、キャッシュメモリ 51 に関する情報を管理するための制御情報と、定義することができる。キャッシュメモリ 51 は、各クラスタ単位で設けられるものであり、その記憶内容はクラスタに依存する。即ち、クラスタが異なれば、キ

キャッシュメモリ 51 に記憶されているデータも相違し得る。そして、ディレクトリ情報の内容は、キャッシュメモリ 51 の記憶内容に依存する。従って、管理情報の場合とは異なり、クラスタを超えてディレクトリ情報を多重化する必要性に乏しい。

【0044】

上記の知見に基づいて、合計 4 個の共有メモリの利用方法を検討する。まず、図 3 (a) に示すように、4 個の共有メモリの全てに、それぞれ管理情報及びディレクトリ情報を記憶させ、共有メモリの構成を全 2 重化することが考えられる。例えば、管理情報は、4 個の共有メモリ全てにおいて同一の内容となるように制御する (4 重化)。また、例えば、ディレクトリ情報は、各クラスタ内において 2 重化する。

【0045】

しかし、この場合は、I/O 処理 (書込み要求処理、読出し要求処理) が発生するたびに、4 個の管理情報の全てをそれぞれ更新させ、各クラスタ内では 2 個のディレクトリ情報をそれぞれ更新させる必要がある。従って、共有メモリの構成を全 2 重化する場合は、制御情報の更新負荷が大きくなる。

【0046】

そこで、次に、図 3 (b) に示すように、制御情報を分割して共有メモリに記憶させることが考えられる。図 3 (b) では、制御情報を「分割情報 1」と「分割情報 2」とに分割している。図 3 (b) の内容をより具体的に示したのが図 4 である。図 4 (a) に示すように、1 つの管理情報を「管理情報 α 」と「管理情報 β 」に分割し、1 つのディレクトリ情報を「ディレクトリ情報 α 」と「ディレクトリ情報 β 」に分割する。そして、例えば、同一クラスタ内の基本メモリ部の共有メモリに「管理情報 α 」及び「ディレクトリ情報 α 」を、増設メモリ部に「管理情報 β 」及び「ディレクトリ情報 β 」を、それぞれ記憶させる。図 4 (b) に、4 個の共有メモリの全体構成を示す。この場合は、冗長性の効果を高くすることができる。しかし、管理情報 α と管理情報 β の双方について常に冗長構成を確保しなければならない。従って、共有メモリのメモリコントロール回路は、管理情報 α 、 β を 2 重化するために、常に 2 重書込み処理を行う必要があり、ディスクアレイ装置全体のさらなる性能向上が望めない。

【0047】

そこで、図 5 に、制御情報を「管理情報」と「ディレクトリ情報」とに分割して複数の共有メモリ 61A ~ 61D にそれぞれ記憶させる様子を模式的に示す。一方のクラスタ 11A では、基本メモリ部の共有メモリ 61A に管理情報 D1 を記憶させ、増設メモリ部の共有メモリ 61B にディレクトリ情報 D2A を記憶させる。同様に、他方のクラスタ 11B では、基本メモリ部の共有メモリ 61C に管理情報 D1 を記憶させ、増設メモリ部の共有メモリ 61D にディレクトリ情報を記憶させる。ここで、各クラスタ 11A、11B がそれぞれ管理する管理情報 D1 は、同一内容である。即ち、管理情報 D1 は、2 重化されている。各クラスタ 11A、11B がそれぞれ管理するディレクトリ情報 D2A、D2B は、各クラスタ 11A、11B に固有の内容である。即ち、ディレクトリ情報 D2A、D2B (以下、区別しない場合は「ディレクトリ情報 D2」と示す) は、2 重化されていない。

【0048】

各共有メモリ 61A、61C では、常に管理情報 D1 の 2 重化を行っているため、メモリコントロール回路の負荷は、図 4 に示す場合と同等である。しかし、ディレクトリ情報 D2A、D2B は 2 重管理していないため、共有メモリ 61B、61D のメモリコントロール回路の負荷を低減することができる。また、情報の属性に基づいて、管理情報 D1 とディレクトリ情報 D2A、D2B とに分割し、それぞれ別々の共有メモリに格納させるため、保守作業等の簡易化を図ることもできる。

【0049】

次に、図 6 は、本実施例による保守制御モードの選択方法の概略を示すフローチャートである。本実施例では、3 種類の保守制御モードの中からいずれか 1 つの保守制御モードを選択することができる。本実施例では、後述する共有メモリのあまり領域の大きさに基

づいて、工場出荷時等に予め保守制御モードが決定される。従って、図6に示すフローチャートは、ディスクアレイ装置10により実行されるものではない。但し、これに限らず、例えば、SVP2を介して保守制御モードを切替可能な構成としてもよい。

【0050】

本実施例では、第1の保守制御モード（「保守制御モード1」と図示）と、第2の保守制御モード（「保守制御モード2」と図示）と、第3の保守制御モード（「保守制御モード3」と図示）との3種類のモードを備えている。第1の保守制御モードは、管理情報D1を記憶する共有メモリ（61A、61C）のあまり領域を利用しないモードである。第2の保守制御モードは、あまり領域のサイズがディレクトリ情報のサイズ以上である場合に、あまり領域を利用してディレクトリ情報の再構築を行うモードである。第3の保守制御モードは、あまり領域のサイズがディレクトリ情報のサイズ未満である場合に、あまり領域を利用してディレクトリ情報の再構築を行うモードである。ここで、あまり領域（空き領域）とは、ディスクアレイ装置10が正常に動作している場合は使用されることはないが、共有メモリに障害が発生した場合には障害回復のために使用される記憶領域を意味する。具体的には、管理情報D1を記憶している記憶領域以外の未使用記憶領域、及び／または、ディレクトリ情報D2を記憶している記憶領域以外の未使用記憶領域が、「あまり領域」に該当する。あまり領域については、さらに後述する。

【0051】

図6に示すように、あまり領域を使用しない場合（S1：NO）、第1の保守制御モードが選択される（S2）。あまり領域を使用する場合において（S1：YES）、あまり領域のサイズがディレクトリ情報のサイズ以上である場合は（S3：YES）、第2の保守制御モードが選択される（S4）。あまり領域のサイズがディレクトリ情報のサイズ未満の場合は（S3：NO）、第3の保守制御モードが選択される。上述の通り、あまり領域のサイズは、共有メモリ61のメモリサイズ等により既に定まっているため、いずれの保守制御モードを使用するかは、予め設定されている。従って、図6に示すフローチャートは、例えば、保守制御モードを選択するための指標として使用され、コンピュータプログラムとして実行されるとは限らない。

【0052】

〔第1の保守制御モード〕

図7～図9に基づき第1の保守制御モードの概要を説明する。図8、図9は、一方のクラスタ11Aを中心に本モードの処理概要の一部を模式的に示す説明図である。他方のクラスタ11Bの場合でも、同様の動作が行われるため、説明を省略する。なお、以下に説明する各保守制御モードは、ディスクアレイ装置10により実行される。より詳しくは、例えば、CHA20、DKA30、SVP2及びメモリコントロール回路62等の協働作業によって、保守制御を実行することができる。但し、これに限らず、保守制御処理用のプロセッサを実装し、この保守制御用プロセッサにより実行させてもよい。

【0053】

まず、ディスクアレイ装置10は、共有メモリ61に障害が発生したか否かを監視している（S11）。例えば、メモリコントロール回路62により共有メモリ61への書込みエラーや読出しエラー等が検出された場合、共有メモリ61に障害が発生したと判断することができる。

【0054】

共有メモリ61の障害発生が検出された場合（S11：YES）、メモリ基本部（Basic）側の共有メモリ61に障害が発生したか否か、即ち管理情報D1に障害が発生したか否かを判定する（S12）。なお、実際には、障害発生を検出した時点で、いずれの共有メモリ61に障害が発生したのか直ちに判明するため、S11とS12とを区別せずに実行可能である。

【0055】

図8（a）にも示すように、基本メモリ部の共有メモリ61（61Aまたは61C）に障害が発生した場合は（S12：YES）、この障害の発生したキャッシュパッケージ（図中「

基板」と略記)の保守閉塞処理を行う(S13)。共有メモリ61Aに障害が発生した場合は、キャッシュパッケージ60Aの保守閉塞が行われ、共有メモリ61Cに障害が発生した場合は、キャッシュパッケージ60Cの保守閉塞が行われる。なお、一方の管理情報D1に障害が発生した場合でも、他方の管理情報D1によってディスクアレイ装置10を正常に動作させることができる。

【0056】

障害の発生したキャッシュパッケージ60の保守閉塞処理を行った後、システム管理者は、ディスクアレイ装置10から障害の発生したキャッシュパッケージ60を取り出し、正常品のキャッシュパッケージ60と交換する。ディスクアレイ装置10は、キャッシュパッケージ60の交換を確認すると(S14: YES)、交換されたキャッシュパッケージ60に搭載されている共有メモリ61に、他面で管理されている管理情報D1をコピーする(S15)。他面で管理されている管理情報とは、障害の発生したクラスタとは別の正常なクラスタで管理されている管理情報を意味する。図8に示す例では、クラスタ11Bの共有メモリ61Cに記憶されている管理情報D1が、交換されたキャッシュパッケージ60Aの共有メモリ61Aにコピーされる。図7のS11~S15及び図8に示すように、第1の保守制御モードにおいて、管理情報D1を記憶する共有メモリ61に障害が発生した場合は、ディレクトリ情報を記憶する共有メモリ61に何らの影響を与えることなく、保守閉塞処理及びデータ回復処理が行われる。

【0057】

一方、増設メモリ部(Option)側の共有メモリ61(61Bまたは61D)に障害が発生した場合、即ちディレクトリ情報D2に障害が発生した場合は(S12: NO)、この障害の発生したディレクトリ情報D2に対応するキャッシュメモリ51を利用できない。従って、この状況下でホストコンピュータ1からデータが書き込まれた場合は、疑似スルー動作が開始される(S16)。上述の通り、疑似スルー動作とは、ホストコンピュータ1からのデータをディスクドライブ41に記憶させた後でホストコンピュータ1に書き込み完了を報告する動作である。従って、疑似スルー動作の期間中、ディスクアレイ装置10の応答性は低下する。

【0058】

ディレクトリ情報D2を記憶する共有メモリ61に障害が発生した場合は、図9(b)にも示すように、管理情報を記憶している基本メモリ部の共有メモリ61に、上書きでディレクトリ情報D2を再構築する(S17)。障害の発生したディレクトリ情報D2に対応するキャッシュメモリ51のディレクトリ構造等を改めて読み出すことにより、ディレクトリ情報D2を得ることができる。この改めて取得されたディレクトリ情報D2は、基本メモリ部の共有メモリ61に上書きで記憶される。従って、上書きされた共有メモリ61では、管理情報D1が失われることになる。図9に示す場合は、共有メモリ61Aの記憶領域に、初期状態で共有メモリ61Bに記憶されているディレクトリ情報D2Aが上書きで保存される。なお、管理情報D1は2重化されているため、ディレクトリ情報D2の上書きによって一方の管理情報D1が失われた場合でも、他方の管理情報D1によりディスクアレイ装置10を正常に稼働させることができる。

【0059】

ディレクトリ情報D2の再構築が完了するまでの間(S18: NO)、疑似スルー動作モードでホストコンピュータ1からの書き込み要求が処理される。ホストコンピュータ1からの書き込み要求をCHA20が受け付けた場合、CHA20は、要求されたデータをDKA30がディスクドライブ41に格納するのを確認した後で、ホストコンピュータ1に書き込み完了を報告する。

【0060】

基本メモリ部の共有メモリ61へのディレクトリ情報D2の再構築が完了すると(S18: YES)、疑似スルー動作モードが停止される(S19)。ディレクトリ情報D2の再構築により、再びキャッシュメモリ51が利用可能となり、ホストコンピュータ1から書き込みを要求されたデータ(ダーティデータ)を基本メモリ部のキャッシュメモリ51及び増

設メモリ部のキャッシュメモリ 51 で、2 重化できるようになるためである。疑似スルー動作モードが解除されると、CHA 20 は、ホストコンピュータ 1 から受信したデータを各キャッシュメモリ 51 (51A と 51B、または 51C と 51D) で 2 重化した時点で、ホストコンピュータ 1 に対し書込み完了を報告する。従って、ディスクアレイ装置 10 の応答性は、正常時の応答性に回復する。

【0061】

障害の発生した増設メモリ部のキャッシュパッケージ 60 の保守閉塞処理が行われた後 (S20)、システム管理者によって、障害の発生したキャッシュパッケージ 60 が正常なキャッシュパッケージ 60 と交換される。ディスクアレイ装置 10 は、障害の発生したキャッシュパッケージ 60 の交換を確認すると (S21: YES)、交換されたキャッシュパッケージ 60 に搭載されている共有メモリ 61 に、ディレクトリ情報 D2 を改めて再構築させる (S22)。基本メモリ部の共有メモリ 61 に再構築されたディレクトリ情報 D2 を、交換された増設メモリ部の共有メモリ 61 にコピーさせる方法も考えられるが、キャッシュメモリ 51 の最新の記憶内容に基づいてディレクトリ情報 D2 を改めて再構築するのが望ましい。交換された正常品のキャッシュパッケージ 60 に搭載された共有メモリ 61 にディレクトリ情報 D2 を再構築する期間中は、S17 により上書きで再構築されたディレクトリ情報 D2 を用いてデータの 2 重化処理等が行われる。

【0062】

交換された増設メモリ部のキャッシュパッケージ 60 の共有メモリ 61 にディレクトリ情報 D2 が再構築されると (S23: YES)、ディスクアレイ装置 10 は、ディレクトリ情報 D2 の参照先を、基本メモリ部の共有メモリ 61 から増設メモリ部の共有メモリ 61 に切り替える (S24)。そして、基本メモリ部の共有メモリ 61 からディレクトリ情報 D2 を消去し (S25)、図 9 (d) にも示すように、他方のクラスタで管理している管理情報 D1 を、基本メモリ部の共有メモリ 61 にコピーする (S26)。

【0063】

以上のように、第 1 の保守制御モードでは、拡張メモリ部の共有メモリ 61 に障害が発生した場合に、基本メモリ部の共有メモリ 61 を作業領域として使用することにより、早期にディレクトリ情報 D2 を再構築する。従って、障害の発生したキャッシュパッケージ 60 が正常品と交換される前に、疑似スルー動作モードを解除することができ、疑似スルー動作モードの期間を短縮することができる。これにより、障害発生時におけるディスクアレイ装置 10 の応答性を改善することができる。

【0064】

〔第 2 の保守制御モード〕

次に、図 10 ~ 図 15 に基づいて、第 2 の保守制御モードの概要を説明する。後述のように、第 2 の保守制御モードでは、最初の保守制御を行った後で、続いて新たな障害が発生した場合の処理も考慮している。

【0065】

図 10 を参照して、共有メモリ 61 のあまり領域について説明する。共有メモリ 61 のメモリサイズや管理情報 D1 のデータサイズ等によっても相違するが、共有メモリ 61 の少なくとも一部に未使用の記憶領域が発生しうる。例えば、クラスタ 11A における基本メモリ部の共有メモリ 61A には、あまり領域 SC が発生する。また、クラスタ 11B における基本メモリ部の共有メモリ 61C にも、あまり領域 SC が発生する。両 SC のサイズは、同一である。第 2 の保守制御モード及び後述する第 3 の保守制御モードでは、このあまり領域 SC を、障害回復に用いる作業領域としてそれぞれ利用する。なお、拡張メモリ部の共有メモリ 61B、61D においても、あまり領域が発生しうる。後述の説明では、基本メモリ部のあまり領域に「SC1」、増設メモリ部のあまり領域に「SC2」の符号を付してそれぞれ区別する場合がある。また、両者を特に区別しない場合等には単に「SC」の符号を付す。

【0066】

図 11 は、第 2 の保守制御モードの処理概要を示すフローチャートである。S31 ~ S

35は、図7と共に述べたS11～S15と同様の処理を行う。即ち、管理情報D1を記憶する基本メモリ部の共有メモリ61に障害が発生した場合（S31：YES、及びS32：YES）、この障害の発生したキャッシュパッケージ60の保守閉塞処理を行った後（S33）、正常なキャッシュパッケージ60に交換する（S34：YES）。そして、正常なキャッシュパッケージ60の共有メモリ61に、他方のクラスタで管理されている管理情報D1をコピーさせる（S35）。

【0067】

一方、図12にも示すように、ディレクトリ情報D2を記憶する拡張メモリ部の共有メモリ61に障害が発生した場合（S32：NO）、ディスクアレイ装置10は、ダーティデータの2重管理を行うことができないため、正常時の動作モードから疑似スルー動作モードに移行する（S36）。そして、図12（b）にも示すように、ディスクアレイ装置10は、基本メモリ部の共有メモリ61が有するあまり領域SC1を利用して、ディレクトリ情報D2を再構築する（S37）。第2の保守制御モードでは、あまり領域SC1のサイズがディレクトリ情報D2のサイズ以上である場合を前提にするため（ $SC1 \geq D2$ ）、基本メモリ部の共有メモリ61が有するあまり領域SC1内に、ディレクトリ情報D2（図中D2A）を再構築可能である。あまり領域SC1にディレクトリ情報D2を再構築した場合は（S38：YES）、この再構築されたディレクトリ情報D2を用いてダーティデータの2重管理が可能となるため、疑似スルー動作モードを終了する（S39）。

【0068】

そして、ディスクアレイ装置10は、障害の発生した拡張メモリ部のキャッシュパッケージ60の保守閉塞処理を行う（S40）。ディスクアレイ装置10は、システム管理者によって、障害の発生したキャッシュパッケージ60が正常なキャッシュパッケージ60に交換されたことを確認した後（S41：YES）、処理を終了する。

【0069】

図12（c）に示すように、基本メモリ部の共有メモリ61のあまり領域SC1を利用してディレクトリ情報D2を再構築した場合は、この共有メモリ61内に管理情報D1とディレクトリ情報D2との両方が格納されることになる。交換された増設メモリ部のキャッシュパッケージ60に搭載された共有メモリ61には、ディレクトリ情報D2が再構築されない。従って、交換された共有メモリ61の記憶領域は、殆どが未使用のあまり領域SC2となる。

【0070】

第2の保守制御モードでは、ディレクトリ情報D2の回復後に、交換された共有メモリ61にディレクトリ情報D2を直ちに再構築することなく、そのまま放置する。ディレクトリ情報D2は、基本メモリ部の共有メモリ61に記憶されているので、障害発生前の初期状態に戻さずにおいても何ら支障ないためである。また、交換された拡張メモリ部の共有メモリ61にディレクトリ情報D2を直ちに再構築して初期状態に戻す場合よりも、保守回復の完了時間を短縮することができる。

【0071】

図12（c）に示す障害回復時の状態において、さらに新たな障害が発生した場合の保守制御方法について、図13～図15に基づき説明する。図14（a）にも示すように、基本メモリ部の共有メモリ61に管理情報D1及びディレクトリ情報D2を記憶させ、拡張メモリ部の共有メモリ61に何も情報が格納されていない場合において、基本メモリ部の共有メモリ61に別の障害が発生したとする（S51：YES、及びS52：YES）。

【0072】

この場合でも、他方のクラスタで管理されている管理情報D1により、ディスクアレイ装置10の稼働を維持することができる。しかし、この場合は、ディレクトリ情報D2も利用できないため、疑似スルー動作モードが開始される（S53）。ディスクアレイ装置10は、疑似スルー動作によってホストコンピュータ1からの書き込み要求を処理しながら、キャッシュメモリ51の記憶構造を取得し、増設メモリ部の共有メモリ61にディレクトリ情報D2を再構築させる（S54、図14（b））。拡張メモリ部の共有メモリ61

にディレクトリ情報D 2が再構築されると(S55: YES)、疑似スルー動作モードが解除される(S 5 6)。ディスクアレイ装置1 0は、障害の発生した基本メモリ部のキャッシュパッケージ6 0を保守閉塞処理し(S 5 7)、システム管理者によって正常なキャッシュパッケージ6 0に交換されるのを待つ(S 5 8)。そして、正常なキャッシュパッケージ6 0に交換された場合(S58: YES)、ディスクアレイ装置1 0は、他方のクラスタで管理されている管理情報D 1を、交換された基本メモリ部の共有メモリ6 1にコピーさせる(S 5 9、図1 4 (c))。これにより、図1 2 (a)と図1 4 (c)とに示すように、2回目の障害から回復した場合、基本メモリ部及び増設メモリ部の構造は、初期状態に戻る。

【0073】

一方、図1 5にも示すように、2回目の障害が拡張メモリ部の共有メモリ6 1に発生した場合(S51: YES、及びS52: NO)、ディスクアレイ装置1 0は、拡張メモリ部のキャッシュパッケージ6 0を保守閉塞処理する(S 6 0、図1 5 (b))。ディレクトリ情報D 2は既に基本メモリ部の共有メモリ6 1に構築されており、障害の発生した拡張メモリ部の共有メモリ6 1には、ディレクトリ情報D 2が記憶されていない。従って、ディスクアレイ装置1 0は、ディレクトリ情報D 2の再構築等を行うことなく、通常の動作を維持しながら、システム管理者によって正常なキャッシュパッケージ6 0に交換されるのを待つ(S 6 1)。正常なキャッシュパッケージ6 0に交換された場合(S61: YES)、障害から回復したクラスタにおける基本メモリ部及び拡張メモリ部の構成は、図1 5 (a), (c)に示すように、2度目の障害発生前と同一となる。

【0074】

[第3の保守制御モード]

図1 6～図2 1に基づいて、第3の保守制御モードを説明する。第3の保守制御モードは、基本メモリ部の共有メモリ6 1が有するあまり領域のサイズが、ディレクトリ情報D 2のサイズに満たない場合を前提とするものである。

【0075】

図1 6は、第3の保守制御モードの処理概要を示すフローチャートである。S 7 1～S 7 5は、図7と共に述べたS 1 1～S 1 5と同様の処理を行う。即ち、管理情報D 1を記憶する基本メモリ部の共有メモリ6 1に障害が発生した場合(S71: YES、及びS72: YES)、この障害の発生したキャッシュパッケージ6 0の保守閉塞処理を行った後(S 7 3)、正常なキャッシュパッケージ6 0に交換する(S74: YES)。そして、正常なキャッシュパッケージ6 0の共有メモリ6 1に、他方のクラスタで管理されている管理情報D 1をコピーさせる(S 7 5)。

【0076】

一方、図1 7にも示すように、拡張メモリ部の共有メモリ6 1に障害が発生した場合(S71: YES、及びS72: NO)、ディレクトリ情報D 2を利用できない場合である。そこで、ディスクアレイ装置1 0は、疑似スルー動作モードに移行する(S 7 6)。次に、ディスクアレイ装置1 0は、基本メモリ部の共有メモリ6 1が有するあまり領域SC 1を利用して、ディレクトリ情報D 2を再構築させる(S 7 7、図1 7 (b))。

【0077】

ここで、基本メモリ部の共有メモリ6 1が有するあまり領域SC 1のサイズは、ディレクトリ情報D 2のサイズ未満であるから(SC 1<D 2)、基本メモリ部の共有メモリ6 1にディレクトリ情報D 2の全体を完全に再構築することはできない。図1 7 (c)に示すように、基本メモリ部の共有メモリ6 1には、あまり領域SC 1のサイズに応じて、ディレクトリ情報の一部である第1の部分ディレクトリ情報D 2 A 1が再構築される。

【0078】

ディスクアレイ装置1 0は、第1の部分ディレクトリ情報D 2 A 1で管理可能な容量だけキャッシュメモリ5 1を使用するように、キャッシュメモリ5 1の使用制限を設定する(S 7 8)。図1 8は、使用制限が設定されたキャッシュメモリ5 1の状態を模式的に示す説明図である。図1 8中では、同一クラスタ内のキャッシュメモリ5 1を一つにまとめ

て表示している。各クラスタ 11A, 11B のキャッシュメモリ 51 のうち、使用可能な容量は、図中仮想線で示すように、第 1 の部分ディレクトリ情報 D2A1 によりサポート可能な範囲に限定される。障害が発生していない正常なクラスタにおいても、キャッシュメモリ 51 に使用制限が設定される。そして、第 1 の部分ディレクトリ情報 D2A1 によりサポート可能な領域内に、ホストコンピュータ 1 から書き込まれたダーティデータが格納されていく (S79)。

【0079】

ディスクアレイ装置 10 は、第 1 の部分ディレクトリ情報 D2A1 の再構築が完了したか否かを判定し (S80)、再構築が完了した場合は (S80: YES)、疑似スルー動作を停止させる (S81)。図 18 と共に説明したように、第 1 の部分ディレクトリ情報 D2A1 でサポート可能な範囲だけ各クラスタ 11A, 11B のキャッシュメモリ 51 を使用するように設定したので、各キャッシュメモリ 51 にダーティデータを 2 重化して保持させることができるためである。

【0080】

ディスクアレイ装置 10 は、障害が発生した拡張メモリ部のキャッシュパッケージ 60 を保守閉塞処理し (S82)、システム管理者によって拡張メモリ部のキャッシュパッケージ 60 が交換されたか否かを判定する (S83)。キャッシュパッケージ 60 が交換された場合 (S83: YES)、ディスクアレイ装置 10 は、この交換された拡張メモリ部の共有メモリ 61 に、第 2 の部分ディレクトリ情報 D2A2 を再構築させる (S84、図 17 (c))。第 2 の部分ディレクトリ情報 D2A2 は、ディレクトリ情報 D2 のうち、第 1 の部分ディレクトリ情報 D2A1 により再構築されなかった残余のディレクトリ情報である ($D2 = D2A1 + D2A2$)。

【0081】

ディスクアレイ装置 10 は、第 2 の部分ディレクトリ情報 D2A2 の再構築が完了した場合 (S85: YES)、キャッシュメモリ 51 への使用制限を解除し (S86)、処理を終了する。第 2 の部分ディレクトリ情報 D2A2 の再構築により、ディレクトリ情報 D2 の全体が復旧するので、各クラスタ 11A, 11B のキャッシュメモリ 51 を制限なく使用可能となるためである。

【0082】

このように、第 3 の保守制御モードでは、基本メモリ部の共有メモリ 61 が有するあまり領域 SC1 を障害回復時の作業領域として利用することにより、あまり領域 SC1 のサイズに応じて、ディレクトリ情報 D2 を部分的に再構築する。部分的に再構築された第 1 の部分ディレクトリ情報 D2A1 により、キャッシュメモリ 51 を部分的に使用することが可能となり、ダーティデータの 2 重化を早期に行うことができる。そして、拡張メモリ部のキャッシュパッケージ 60 が交換された場合は、残りのディレクトリ情報 D2A2 を新しい共有メモリ 61 に再構築する。従って、図 17 (c) に示すように、交換された共有メモリ 61 には、第 2 の部分ディレクトリ情報 D2A2 が記憶され、その他の記憶領域はあまり領域 SC2 となる。第 3 の保守制御モードは、各共有メモリ 61 の構成を、1 回目の障害から回復した時点の構成に保持し、初期状態には戻さない。

【0083】

図 17 (c) に示す最初の障害回復状態において、さらに別の障害が発生した場合の処理を、図 19 ~ 図 21 に基づいて説明する。

図 20 (a) にも示すように、新たな障害が基本メモリ部の共有メモリ 61 に発生すると (S91: YES、及び S92: YES)、ディスクアレイ装置 10 は、疑似スルー動作モードに移行する (S93)。基本メモリ部の共有メモリ 61 に部分的に再構築された第 1 の部分ディレクトリ情報 D2A1 を使用できず、各キャッシュメモリ 51 でダーティデータの 2 重化を行うことができないためである。

【0084】

そして、図 20 (b) にも示すように、ディスクアレイ装置 10 は、拡張メモリ部の共有メモリ 61 にディレクトリ情報 D2 全体の再構築を行う (S94)。ディスクアレイ装

置 10 は、ディレクトリ情報 D2 の再構築が完了すると (S95: YES)、疑似スルー動作モードを停止させ (S96)、障害の発生した基本メモリ部のキャッシュパッケージ 60 を保守閉塞させる (S97)。ディスクアレイ装置 10 は、システム管理者により基本メモリ部のキャッシュパッケージ 60 が正常品に交換されたことを確認すると (S98: YES)、他方のクラスタで管理されている管理情報 D1 を、交換された基本メモリ部の共有メモリ 61 にコピーさせる (S99、図 20 (c))。図 17 (a) と図 20 (c) とに示すように、2 度目の障害から回復することにより、各共有メモリ 61 の構成は初期状態に復帰することになる。

【0085】

一方、図 21 (a) にも示すように、拡張メモリ部の共有メモリ 61 に障害が発生した場合 (S91: YES、及び S92: NO)、ディスクアレイ装置 10 は、拡張メモリ部のキャッシュパッケージ 60 を保守閉塞処理させる (S100)。次に、ディスクアレイ装置 10 は、図 18 と共に説明したように、各キャッシュメモリ 51 の使用可能範囲を、第 1 の部分ディレクトリ情報 D2 A1 によりサポート可能な範囲内に制限する (S101)。これにより、疑似スルー動作を行わずに、ホストコンピュータ 1 からの書込み要求を処理することができる。

【0086】

ディスクアレイ装置 10 は、拡張メモリ部のキャッシュパッケージ 60 が正常品と交換されたことを確認すると (S102: YES)、図 16 の S84 ~ S86 で述べたと同様に、交換された共有メモリ 61 に第 2 の部分ディレクトリ情報 D2 A2 を再構築させる (S103、図 21 (c))。そして、ディスクアレイ装置 10 は、第 2 の部分ディレクトリ情報 D2 A2 の再構築が完了した場合 (S104: YES)、各キャッシュメモリ 51 への使用制限を解除する (S105)。

【0087】

このように構成される本実施例によれば、キャッシュメモリ 51 と共有メモリ 61 とをそれぞれ別々のパッケージ 50、60 に分割するため、キャッシュメモリ 51 または共有メモリ 61 のいずれかに障害が発生した場合でも、それぞれ別々に保守回復作業を行うことができる。また、各メモリ 51、61 毎にパッケージを分割し、各パッケージにそれぞれメモリコントロール回路を実装するため、各メモリ 51、61 への並列アクセスを行うことができ、ディスクアレイ装置 10 の高速化に寄与する。

【0088】

さらに、制御情報を管理情報 D1 とディレクトリ情報 D2 とに分割して、別々の共有メモリ 61 に記憶させるため、管理情報 D1 の 2 重化と、ディレクトリ情報 D2 の 1 重化とを効率的に実現できる。従って、ディレクトリ情報 D2 を記憶する拡張メモリ部の共有メモリ 61 の書込み頻度を低下させることができ、ディスクアレイ装置 10 の負荷を低減することができる。換言すれば、制御情報を、多重管理上の要請に基づいて 2 種類の情報 D1、D2 に分割するため、情報管理が容易となり、保守回復の作業性も向上する。

【0089】

さらに、管理情報 D1 とディレクトリ情報 D2 とに分割することにより、ディレクトリ情報 D2 のみを再構築した時点で、疑似スルー動作モードを解除できる。従って、保守回復の全期間にわたって疑似スルー動作を行う必要がなく、疑似スルー動作の実行期間をディレクトリ情報 D2 の再構築に要する時間のみに抑制することができる。これにより、障害発生時におけるディスクアレイ装置 10 の性能低下時間を短縮することができる。

【0090】

また、2 重化されている管理情報 D1 を記憶する基本メモリ部の共有メモリ 61 が有する記憶領域を作業領域として利用するため、障害復旧専用の作業メモリを設けることなく、比較的短時間でディレクトリ情報 D2 の再構築を行うことができる。

【0091】

なお、本発明は、上述した実施の形態に限定されない。当業者であれば、本発明の範囲内で、種々の追加や変更等を行うことができる。例えば、実施例では、各クラスタにおい

て、それぞれ2個ずつのキャッシュメモリ及び共有メモリを搭載する例を示したが、これに限らず、それぞれ3個ずつ以上のメモリから構成することもでき、あるいは、キャッシュメモリと共有メモリとの構成個数を違えることもできる。

【図面の簡単な説明】

【0092】

【図1】本発明の実施例に係わるディスクアレイ装置の全体概要を示すブロック図である。

【図2】キャッシュパッケージの概略構造を示すブロック図である。

【図3】複数の共有メモリに制御情報を記憶させる一例であって、(a)は全ての共有メモリに同一の情報を保持させる場合を、(b)は制御情報を分割して複数の共有メモリに保持させる場合を、それぞれ示す説明図である。

【図4】図3(b)に示す分割例をより具体化して示す説明図である。

【図5】制御情報を管理情報とディレクトリ情報とに分けて共有メモリに記憶させる様子を示す説明図である。

【図6】あまり領域のサイズに応じて保守制御モードを設定可能であることを示すフローチャートである。

【図7】第1の保守制御モードによる処理を示すフローチャートである。

【図8】第1の保守制御モードにおいて、基本メモリ部に障害が発生した場合を模式的に示す説明図である。

【図9】第1の保守制御モードにおいて、増設メモリ部に障害が発生した場合を模式的に示す説明図である。

【図10】管理情報、ディレクトリ情報及びあまり領域の関係を示す説明図である。

【図11】第2の保守制御モードによる処理を示すフローチャートである。

【図12】第2の保守制御モードにおいて、増設メモリ部に障害が発生した場合を模式的に示す説明図である。

【図13】第2の保守制御モードにおいて、再度障害が発生した場合の処理を示すフローチャートである。

【図14】第2の保守制御モードにおいて、基本メモリ部に再度障害が発生した場合を模式的に示す説明図である。

【図15】第2の保守制御モードにおいて、増設メモリ部に再度障害が発生した場合を模式的に示す説明図である。

【図16】第3の保守制御モードによる処理を示すフローチャートである。

【図17】第3の保守制御モードにおいて、増設メモリ部に障害が発生した場合を模式的に示す説明図である。

【図18】部分的に再構築されたディレクトリ情報によって、各クラスタのキャッシュメモリの使用可能範囲が制限される様子を示す説明図である。

【図19】第3の保守制御モードにおいて、再度障害が発生した場合の処理を示すフローチャートである。

【図20】第3の保守制御モードにおいて、基本メモリ部に再度障害が発生した場合を模式的に示す説明図である。

【図21】第3の保守制御モードにおいて、増設メモリ部に再度障害が発生した場合を模式的に示す説明図である。

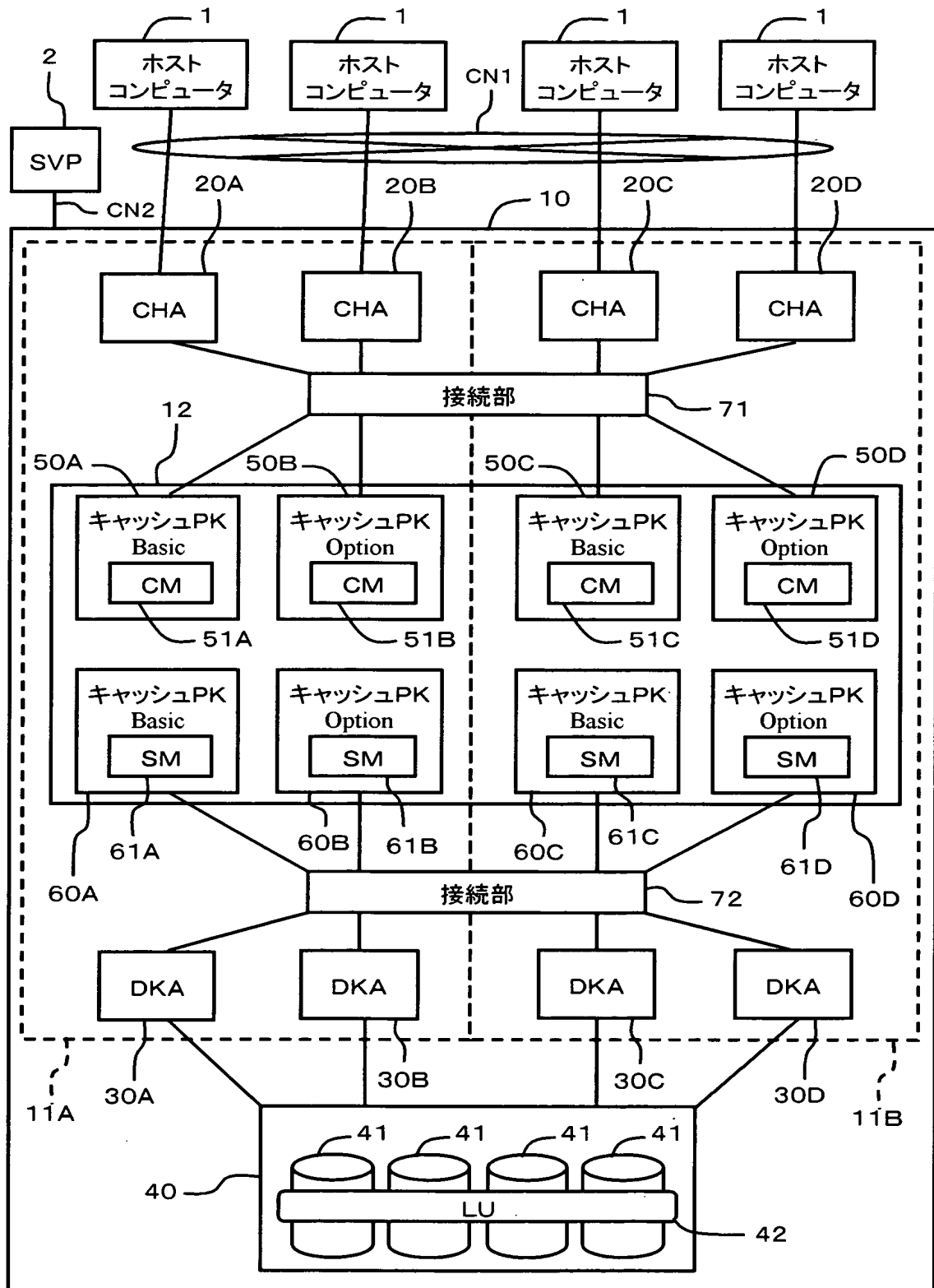
【符号の説明】

【0093】

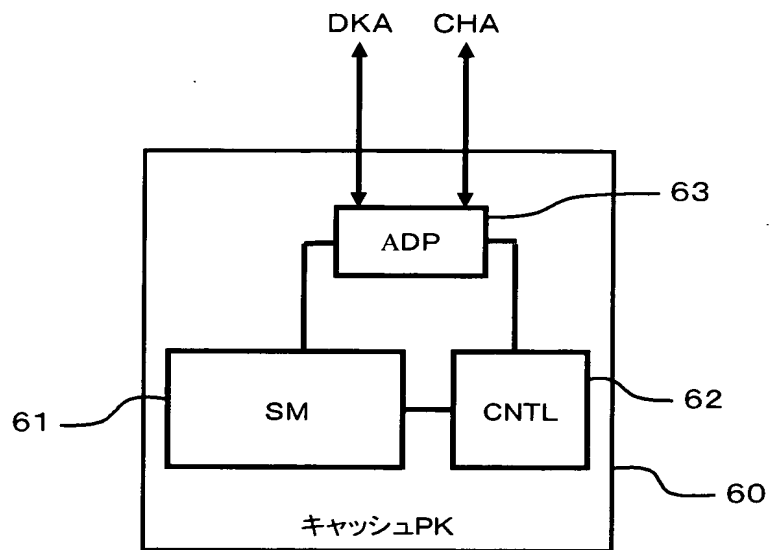
1…ホストコンピュータ、2…SVP、10…ディスクアレイ装置、11A、11B…クラスタ、12…メモリ部、20A～20D…CHA、30A～30D…DKA、40…記憶装置、41…ディスクドライブ、42…論理ボリューム、50A～50D…キャッシュパッケージ(キャッシュメモリ用)、51A～51D…キャッシュメモリ、60A～60D…キャッシュパッケージ(共有メモリ用)、61A～61D…共有メモリ、62…メモリコントロール回路、63…アダプタ、71、72…接続部、D1…管理情報、D2

…ディレクトリ情報、D 2 A, D 2 B…ディレクトリ情報、D 2 A 1, D 2 A 2…部分ディレクトリ情報、S C…あまり領域

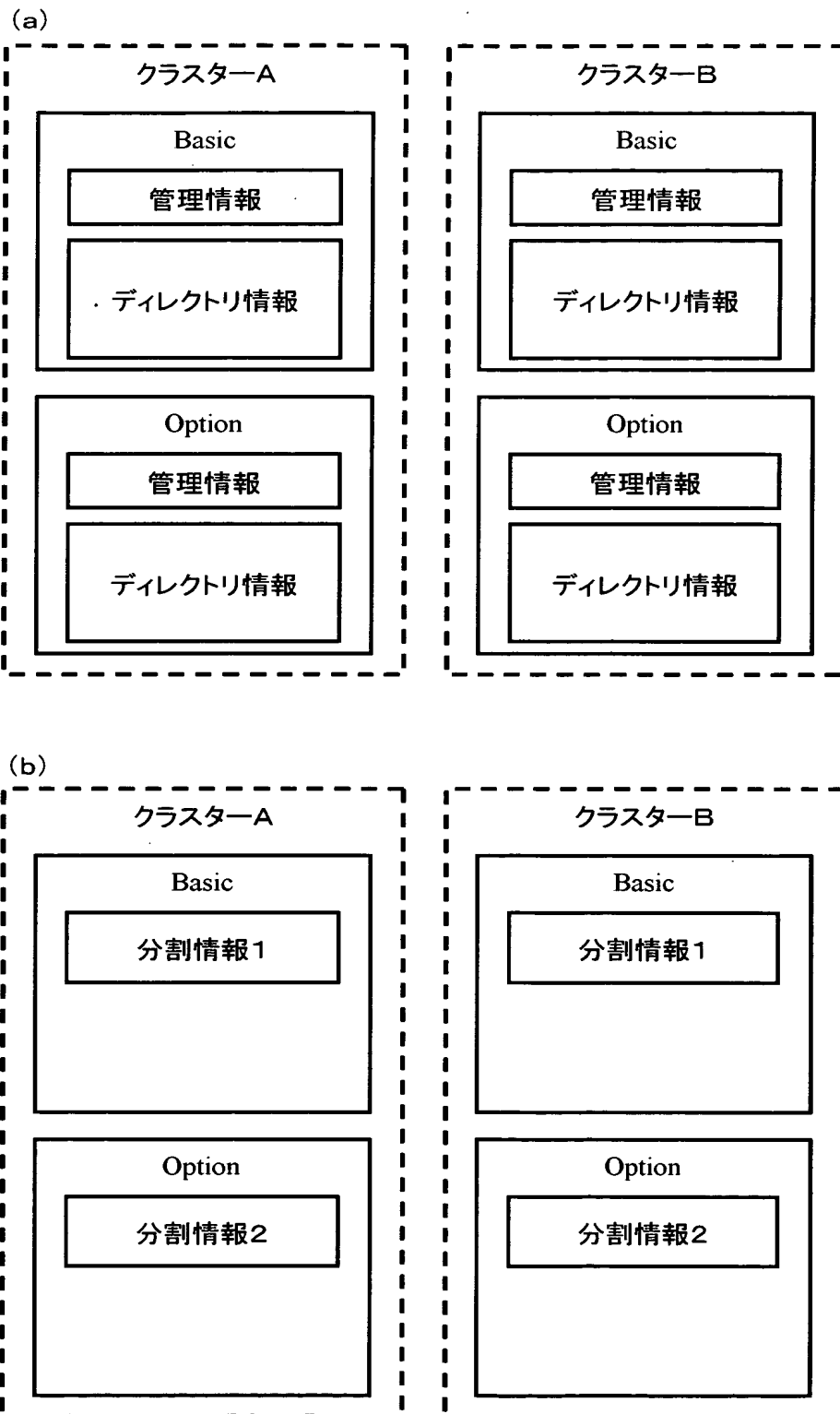
【書類名】 図面
【図 1】



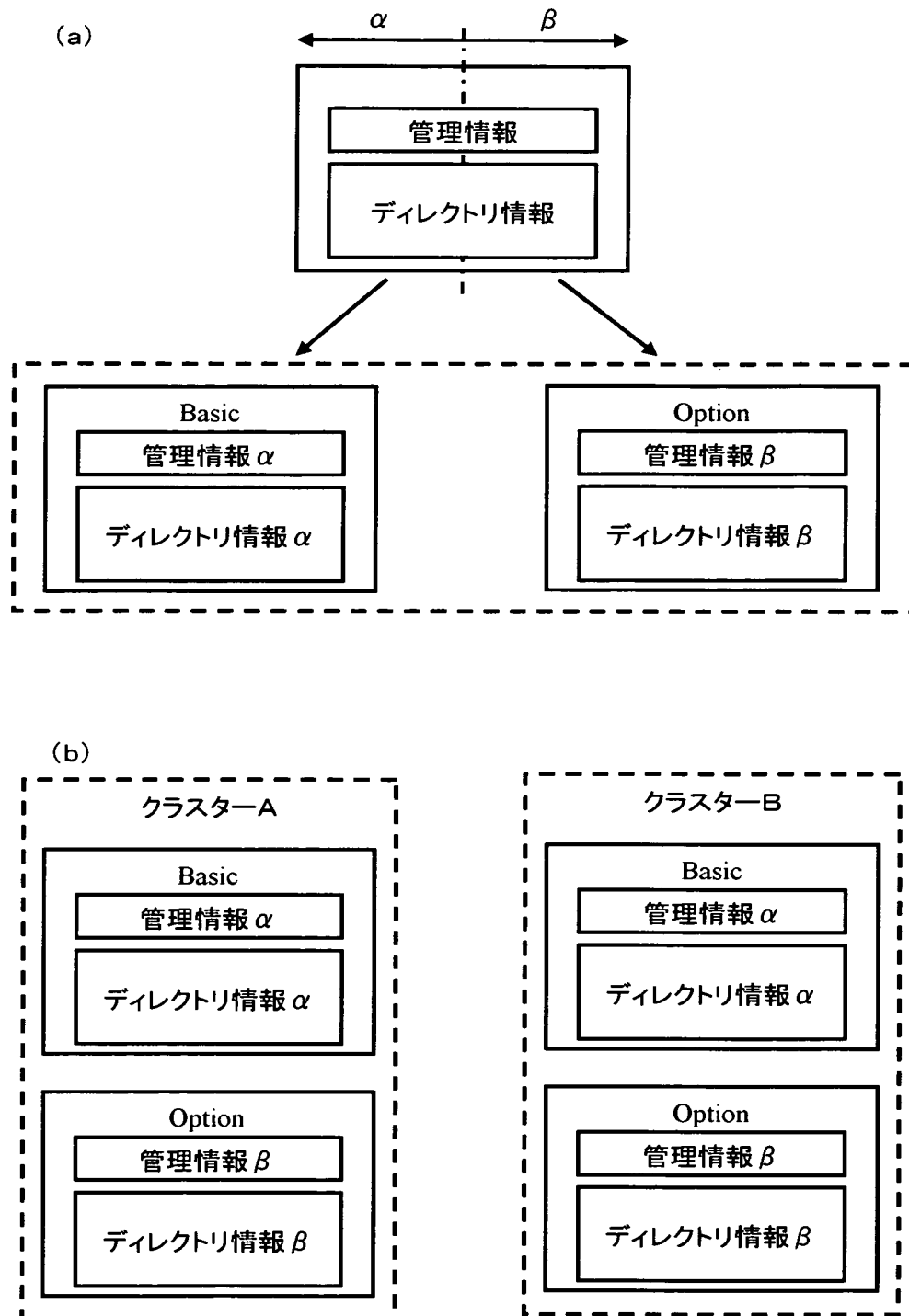
【図 2】



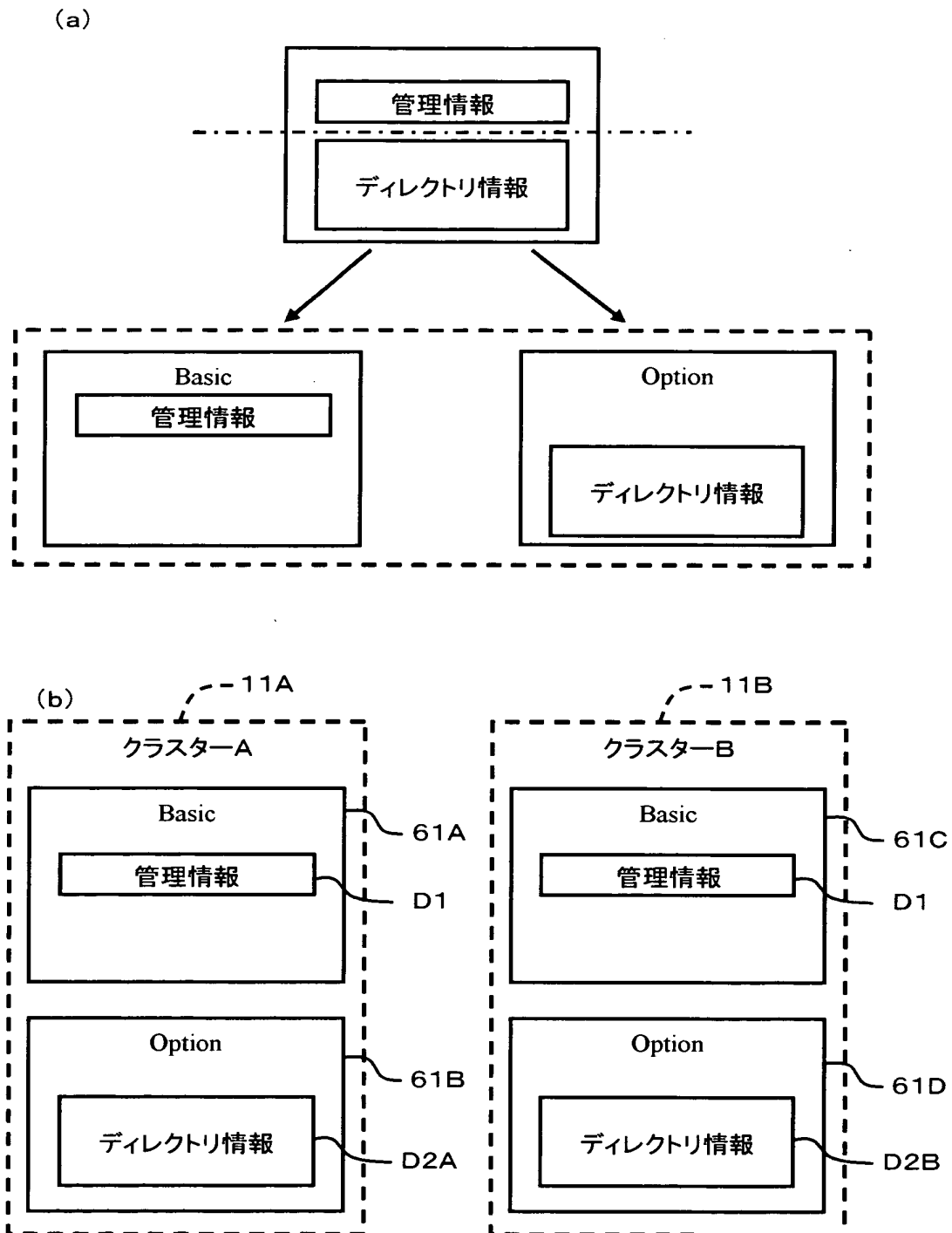
【図 3】



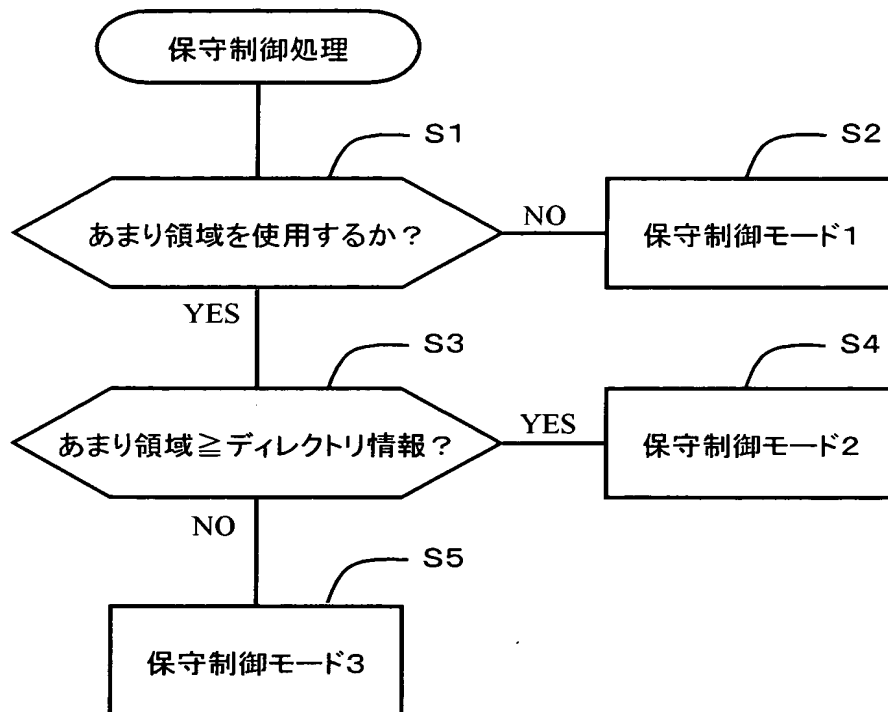
【図 4】



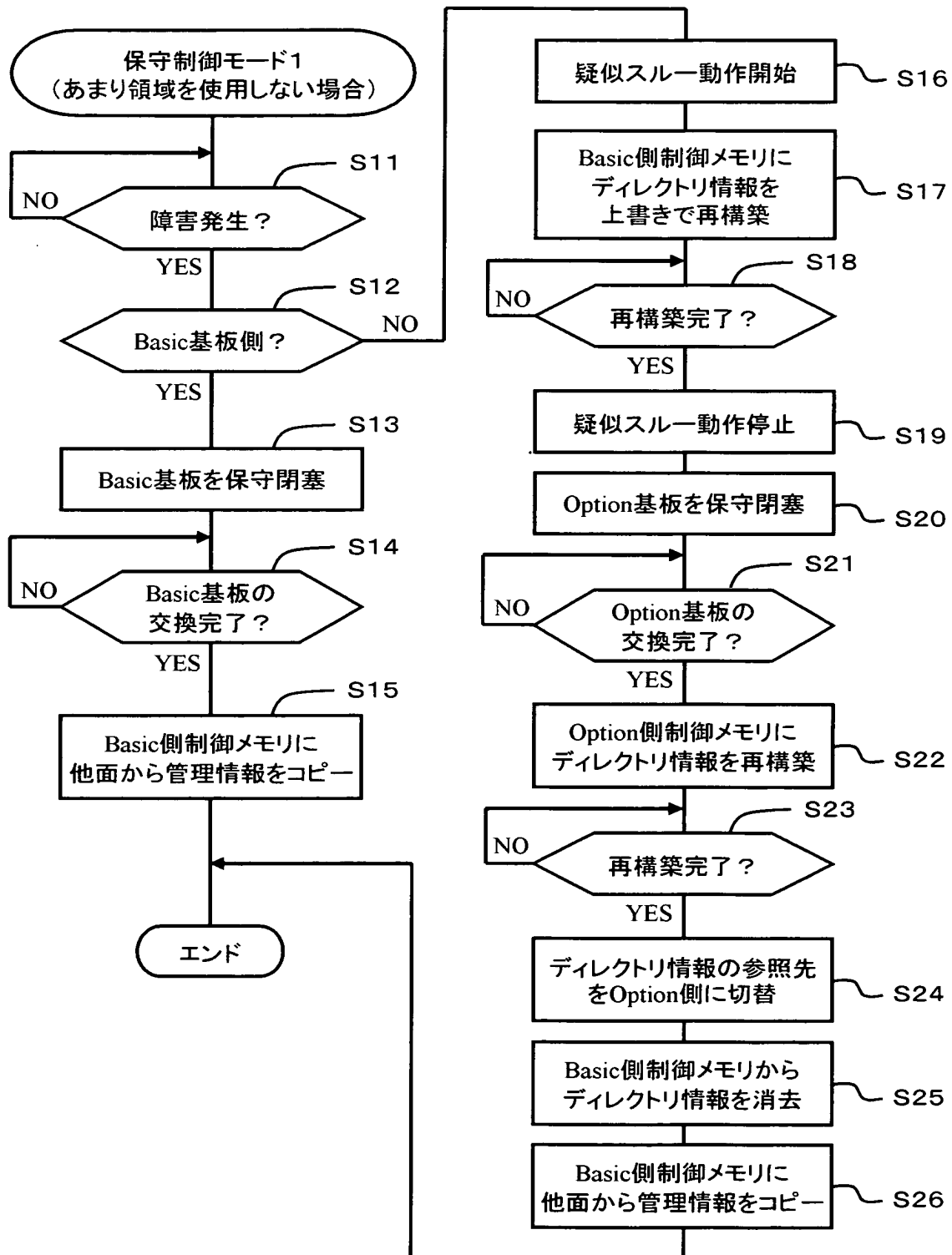
【図 5】



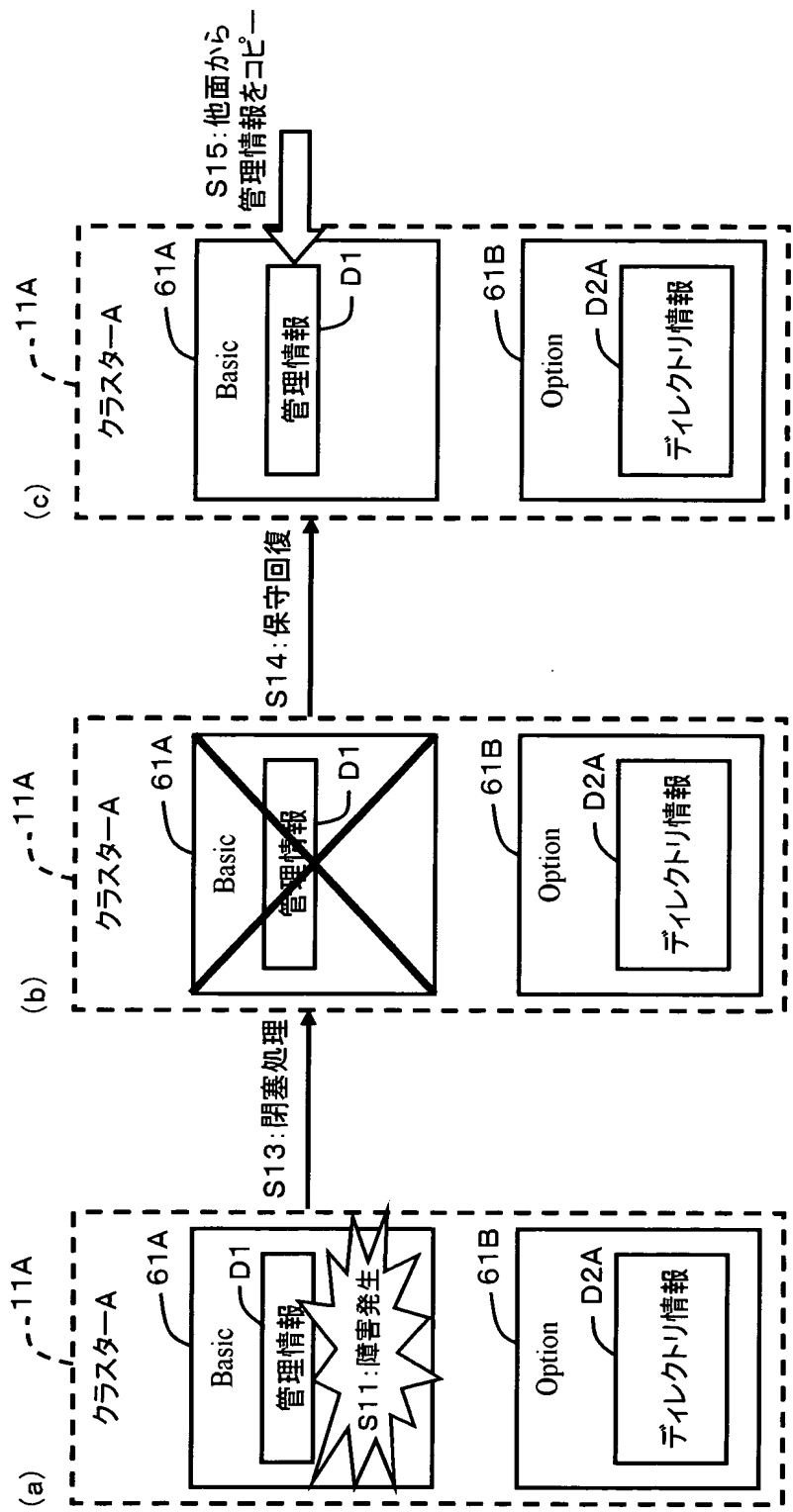
【図 6】



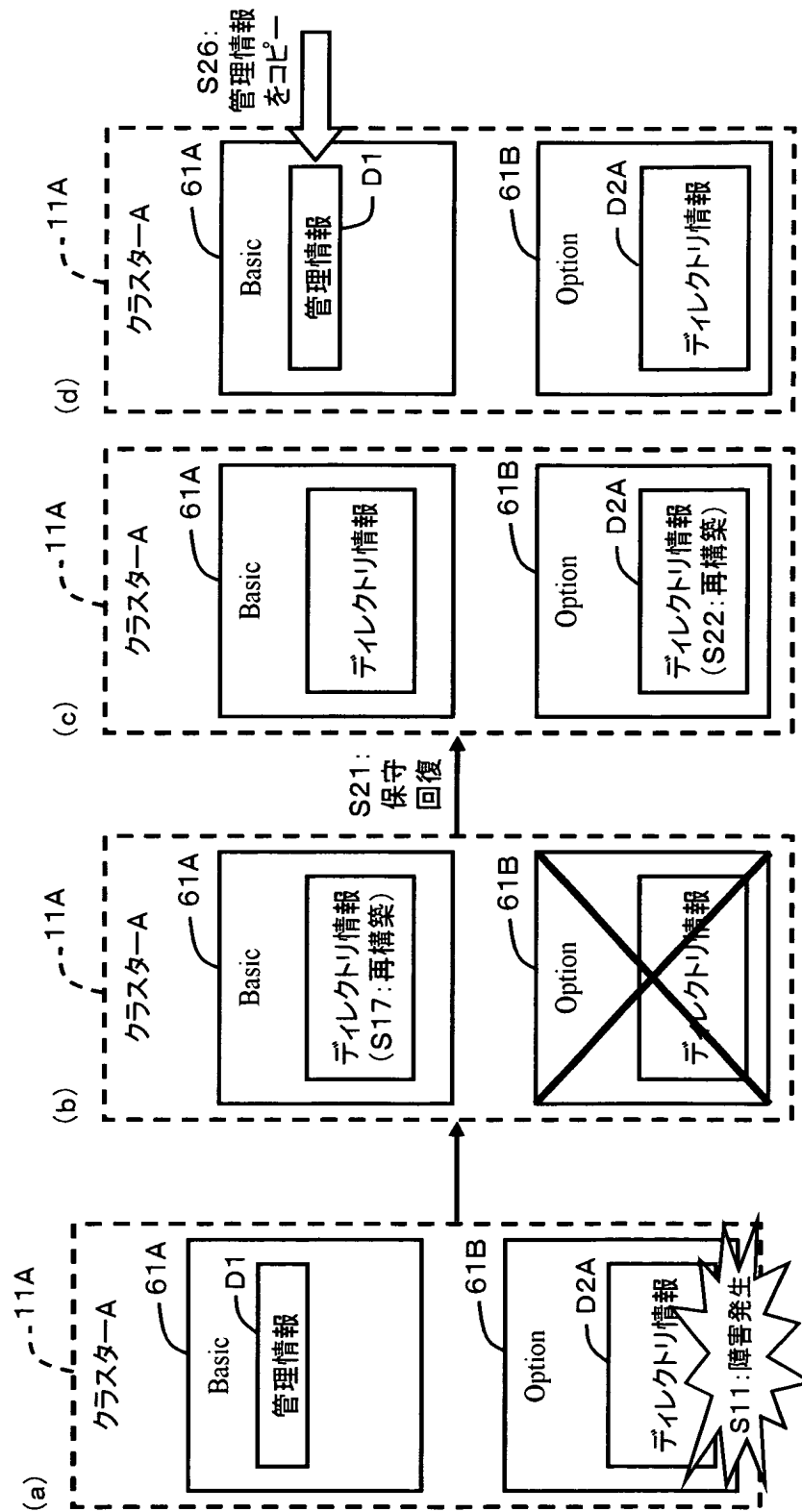
【図 7】



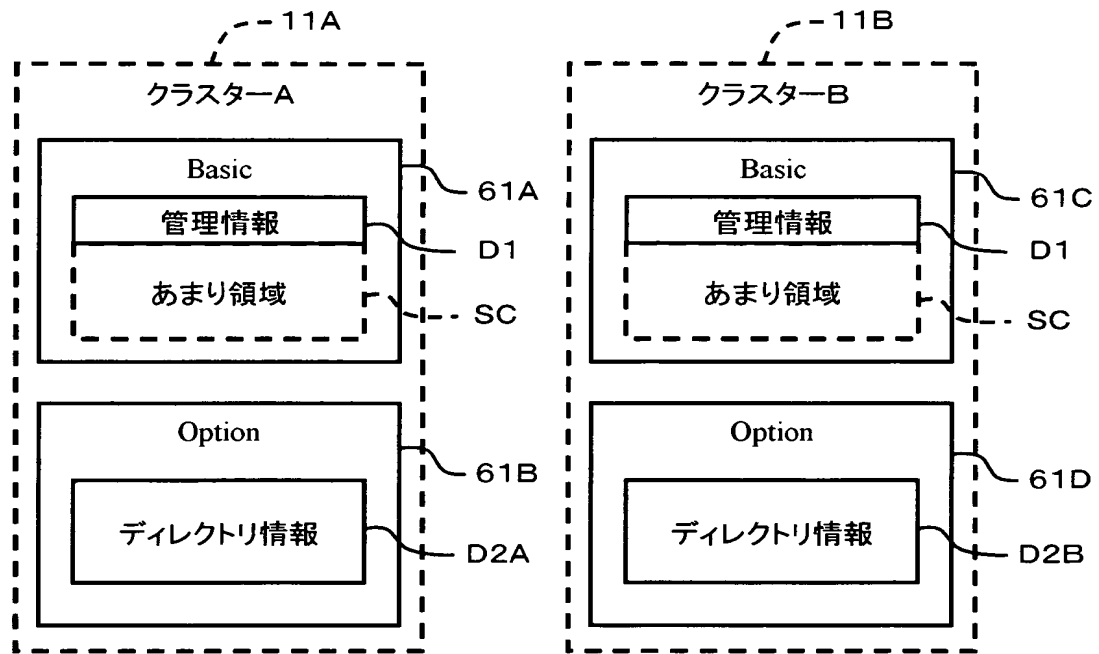
【図 8】



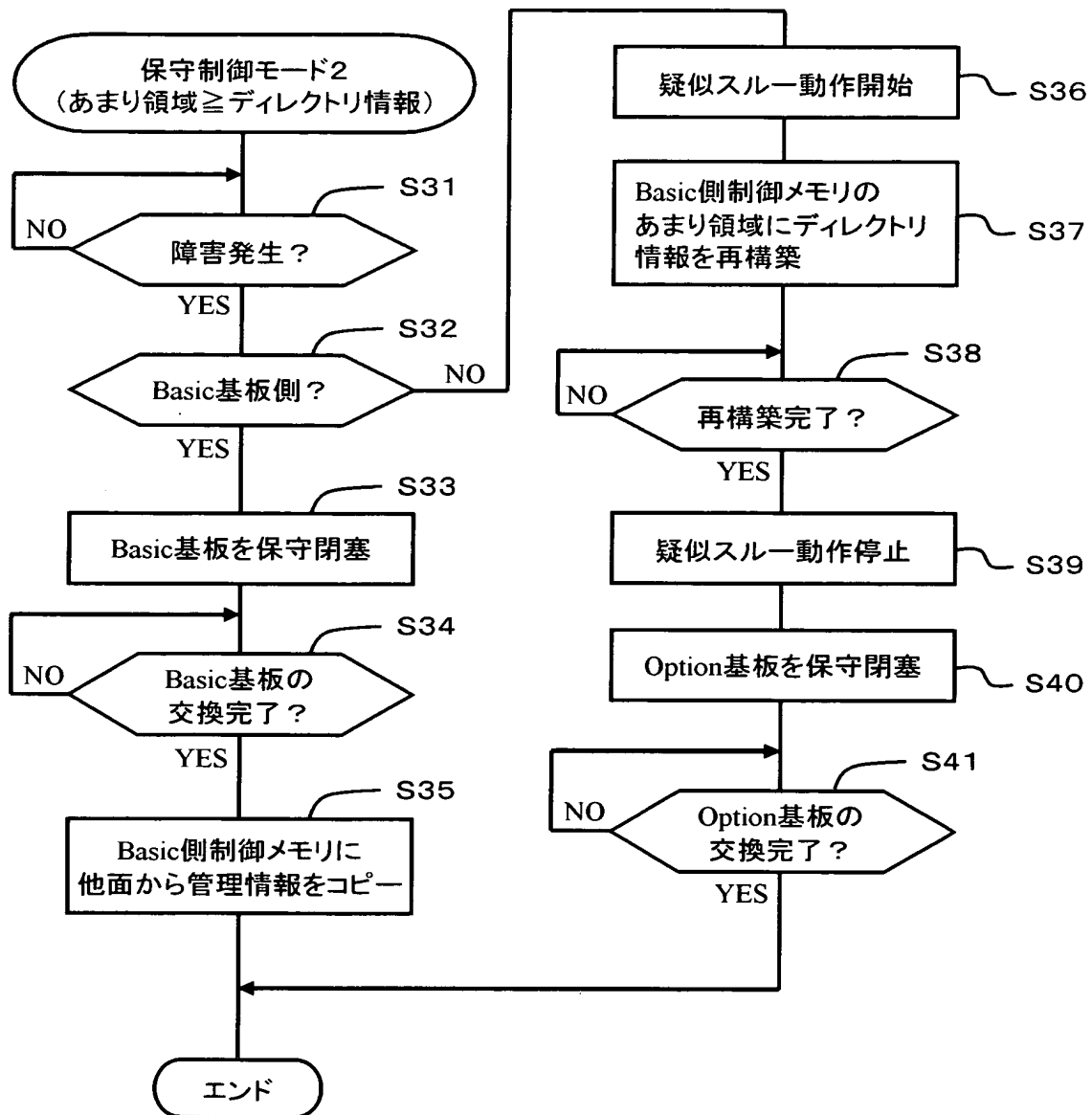
【図 9】



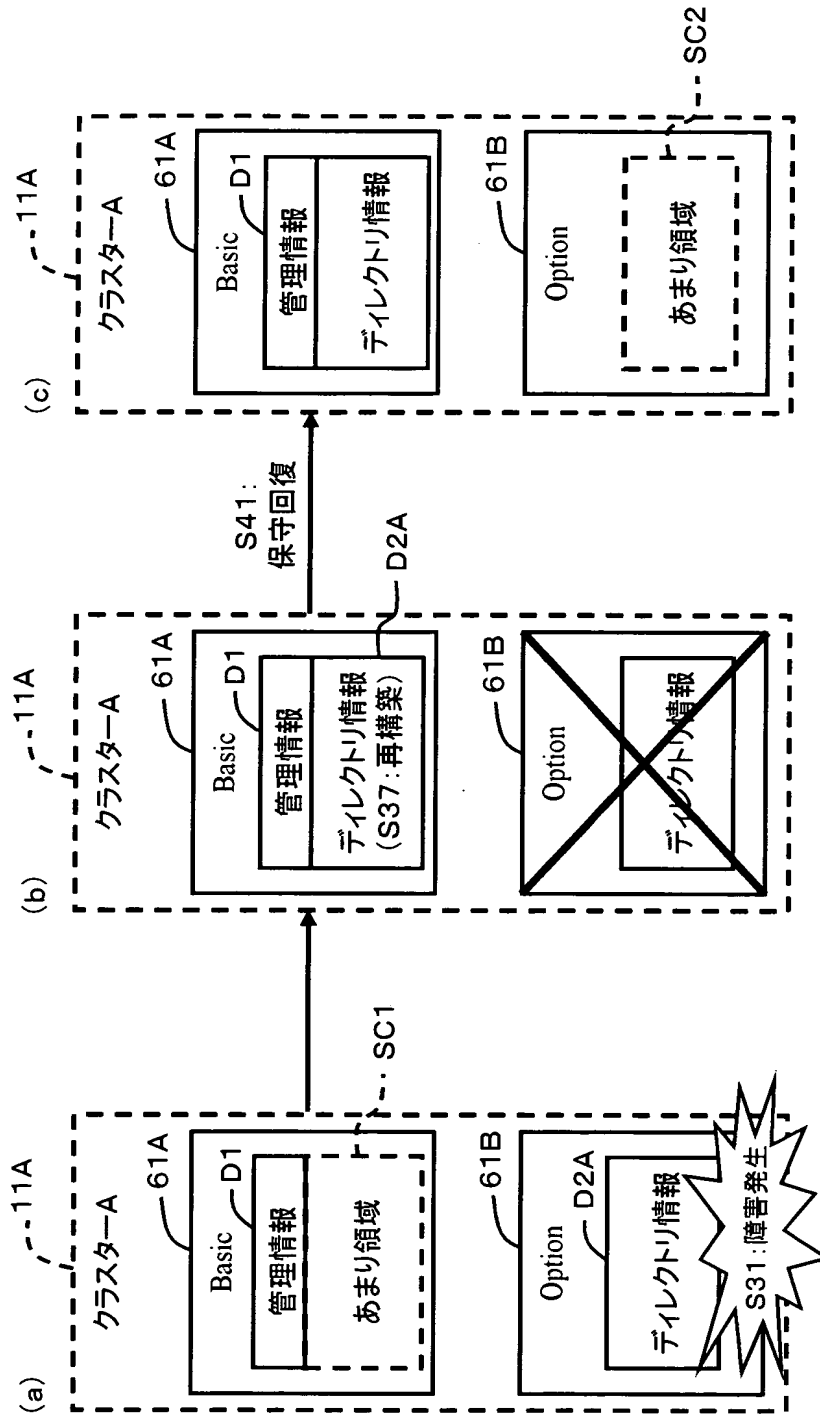
【図 10】



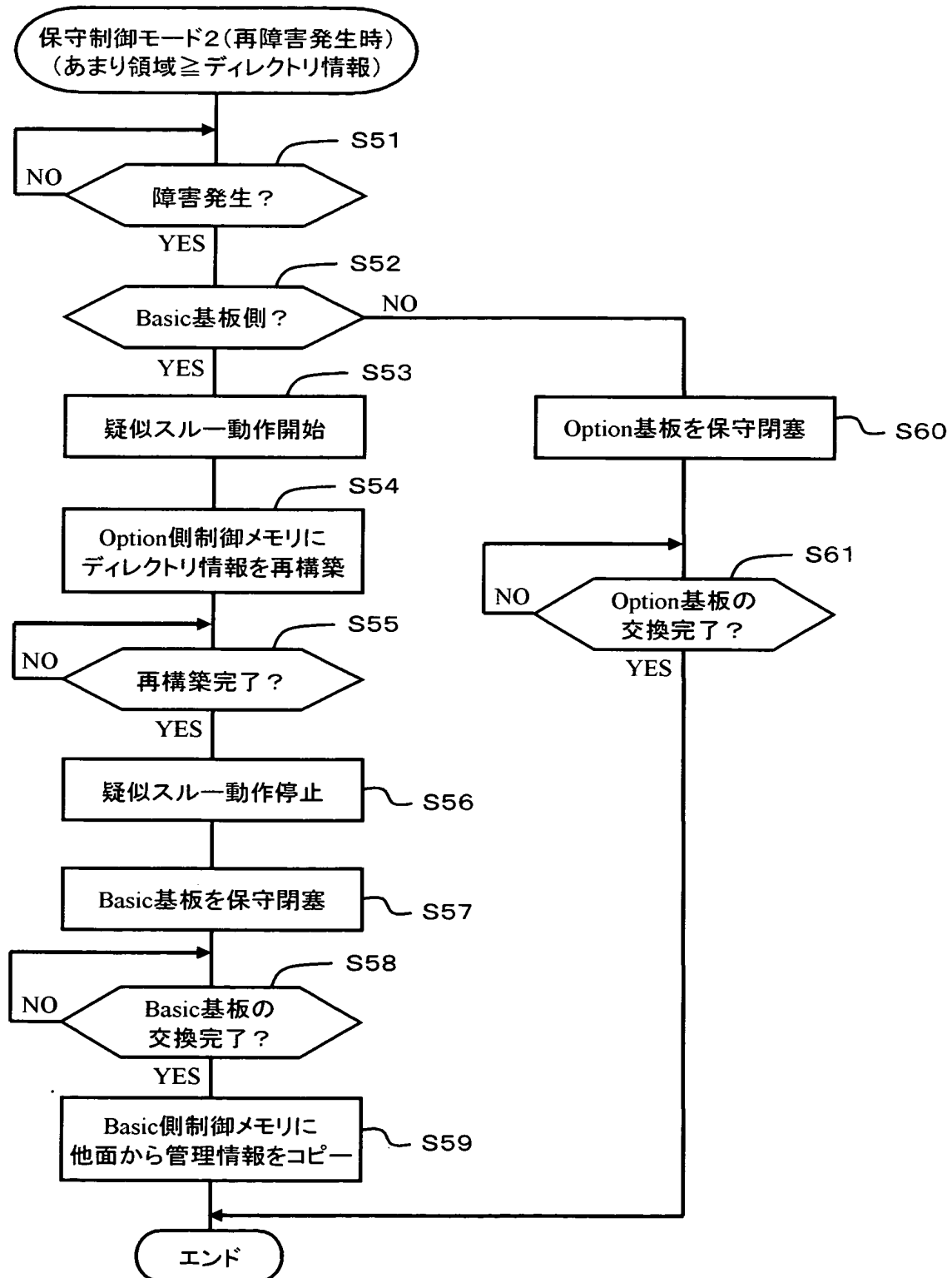
【図 11】



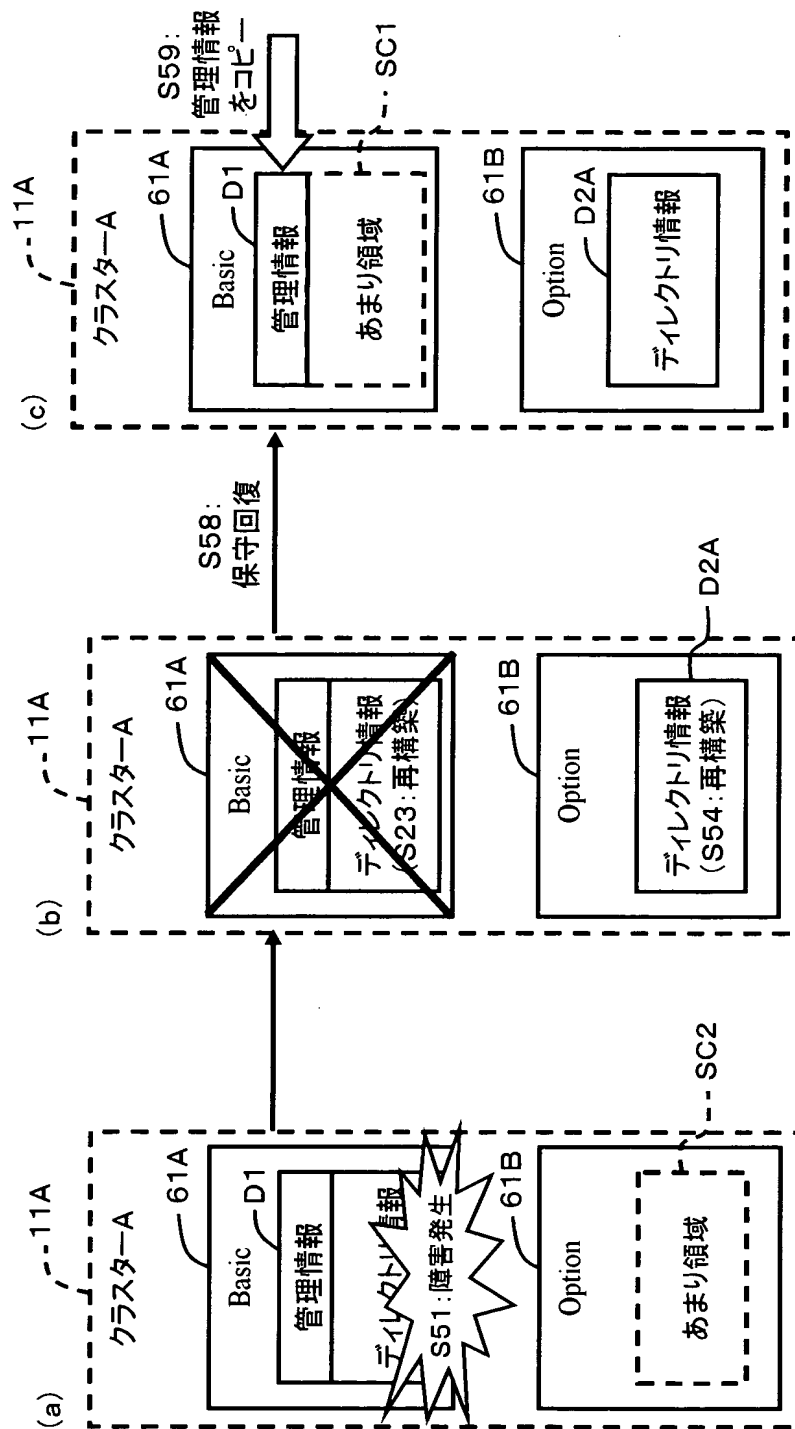
【図 12】



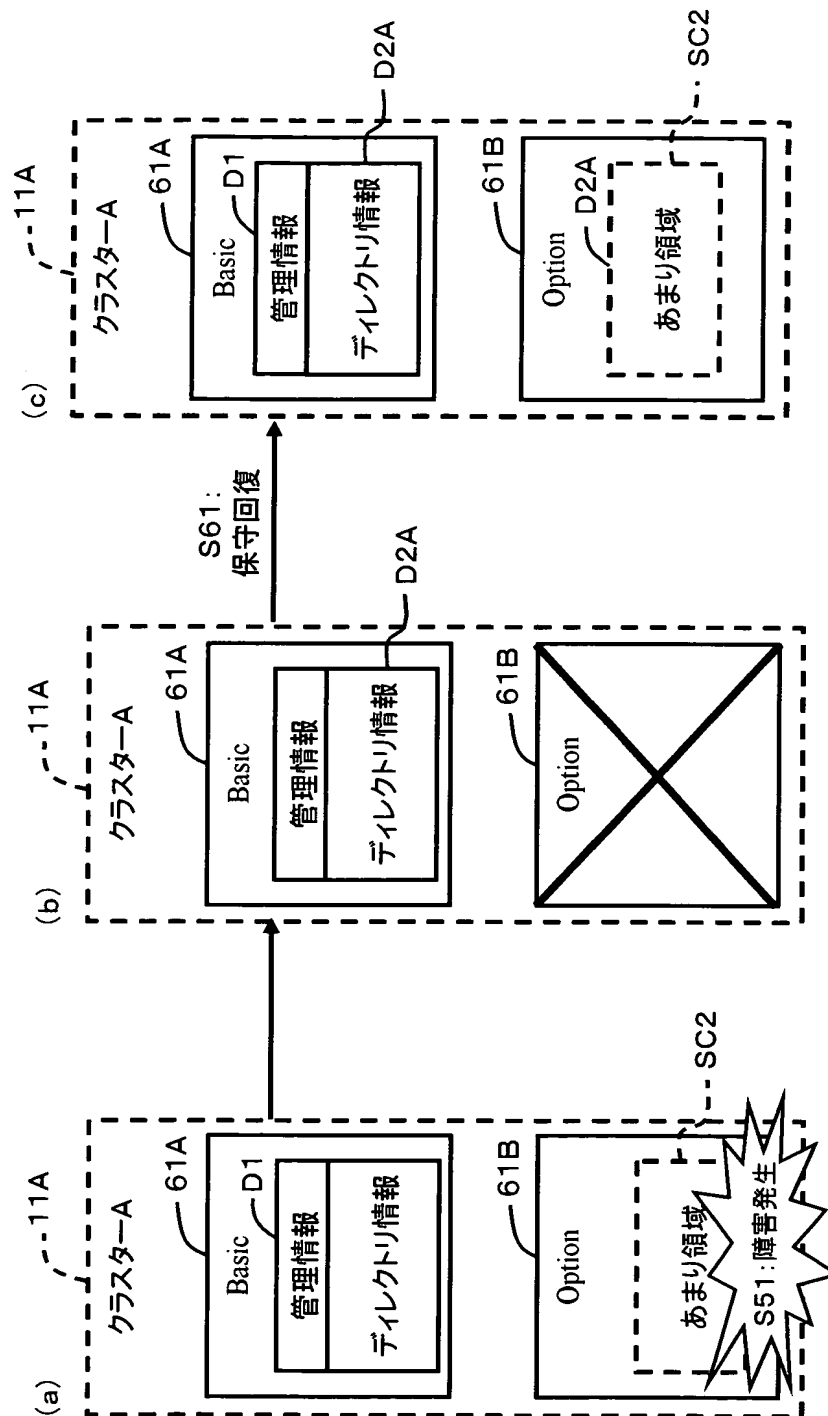
【図 13】



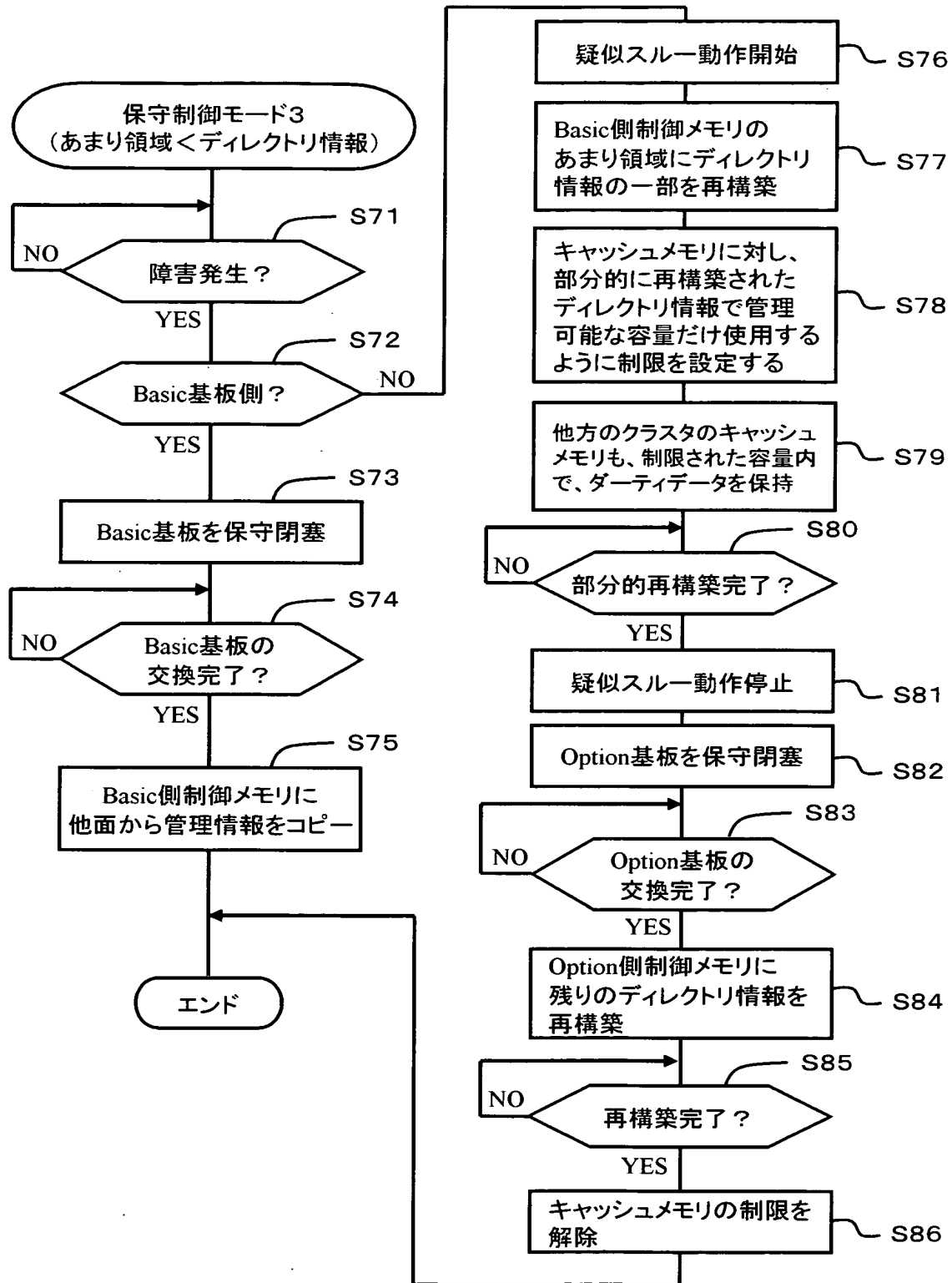
【図 14】



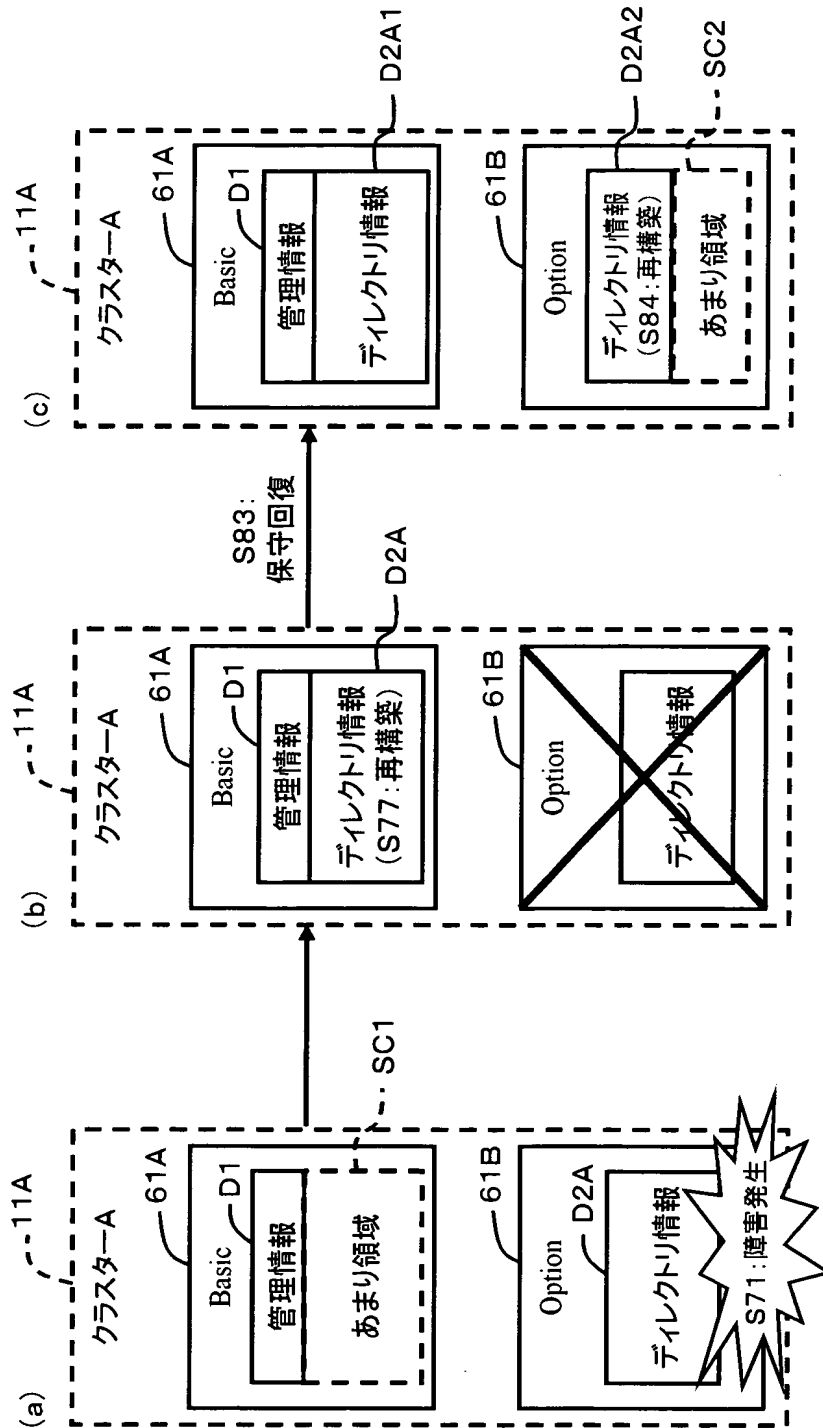
【図 15】



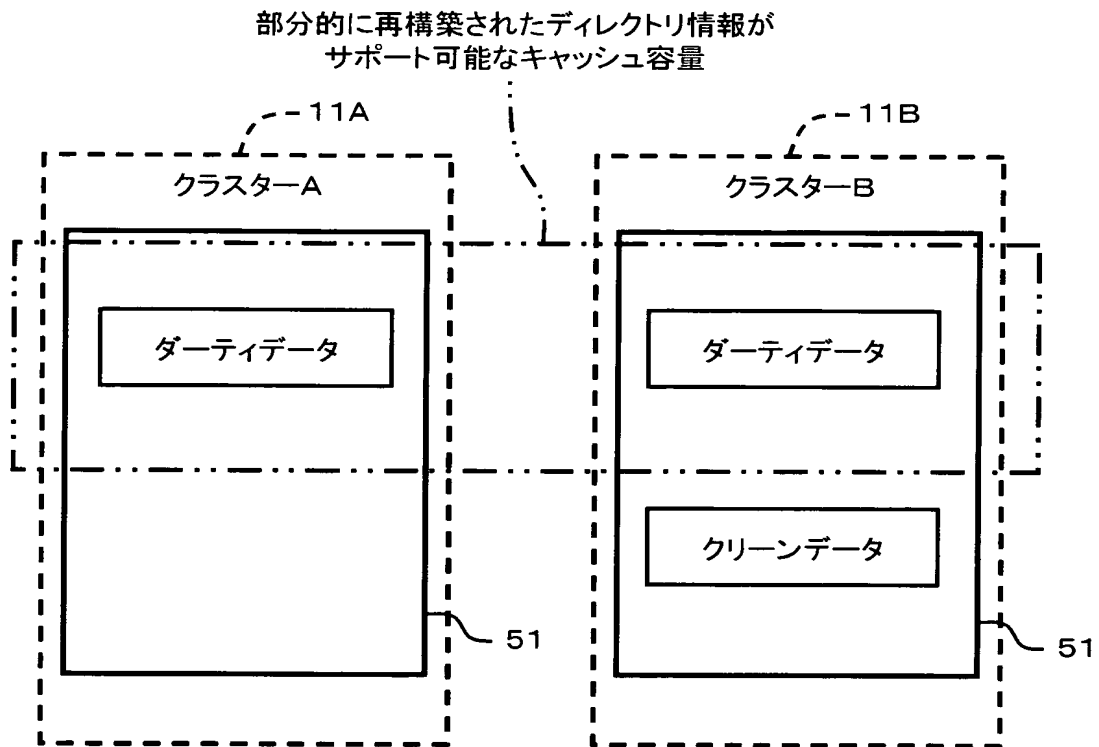
【図 16】



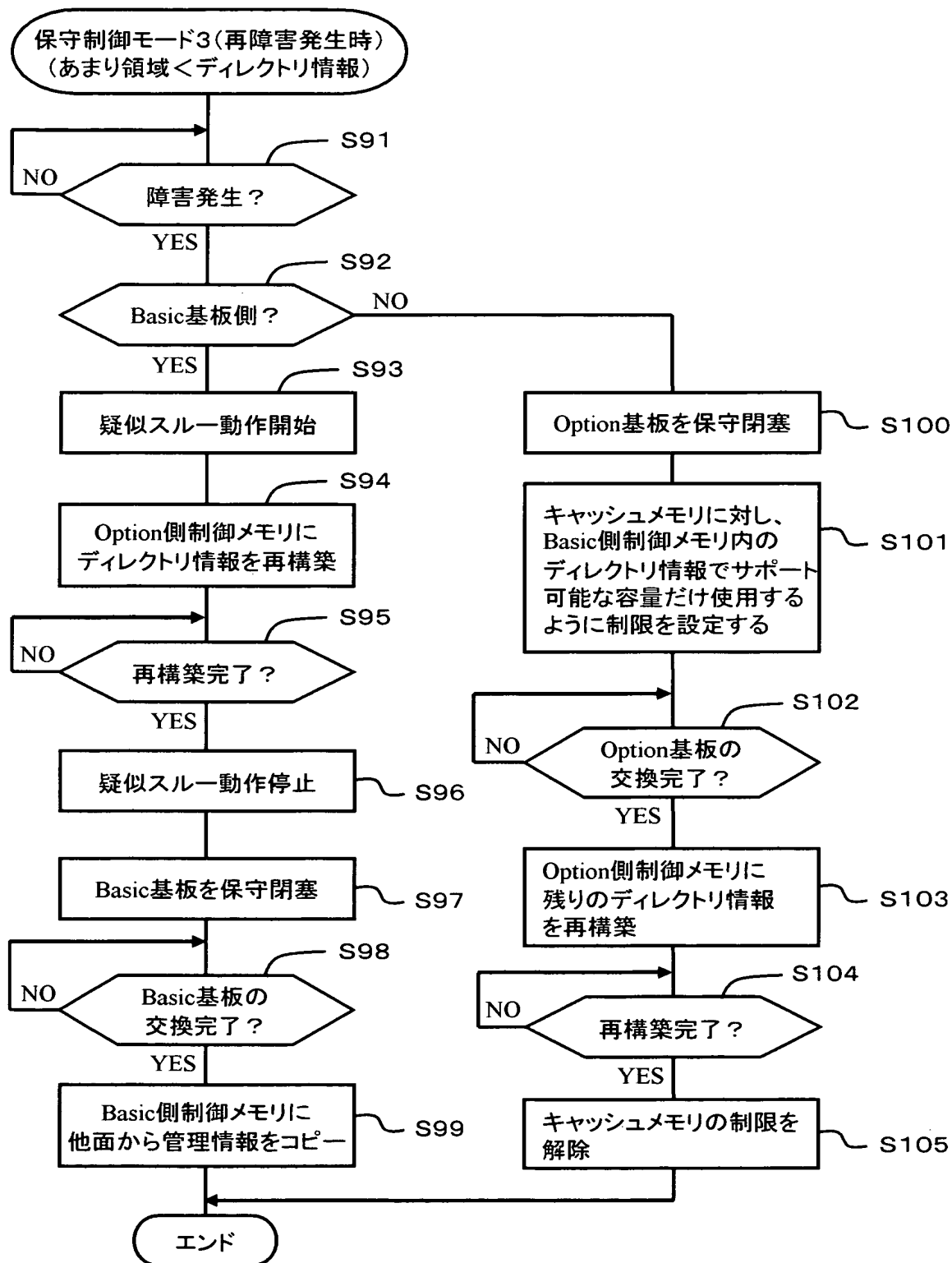
【図 17】



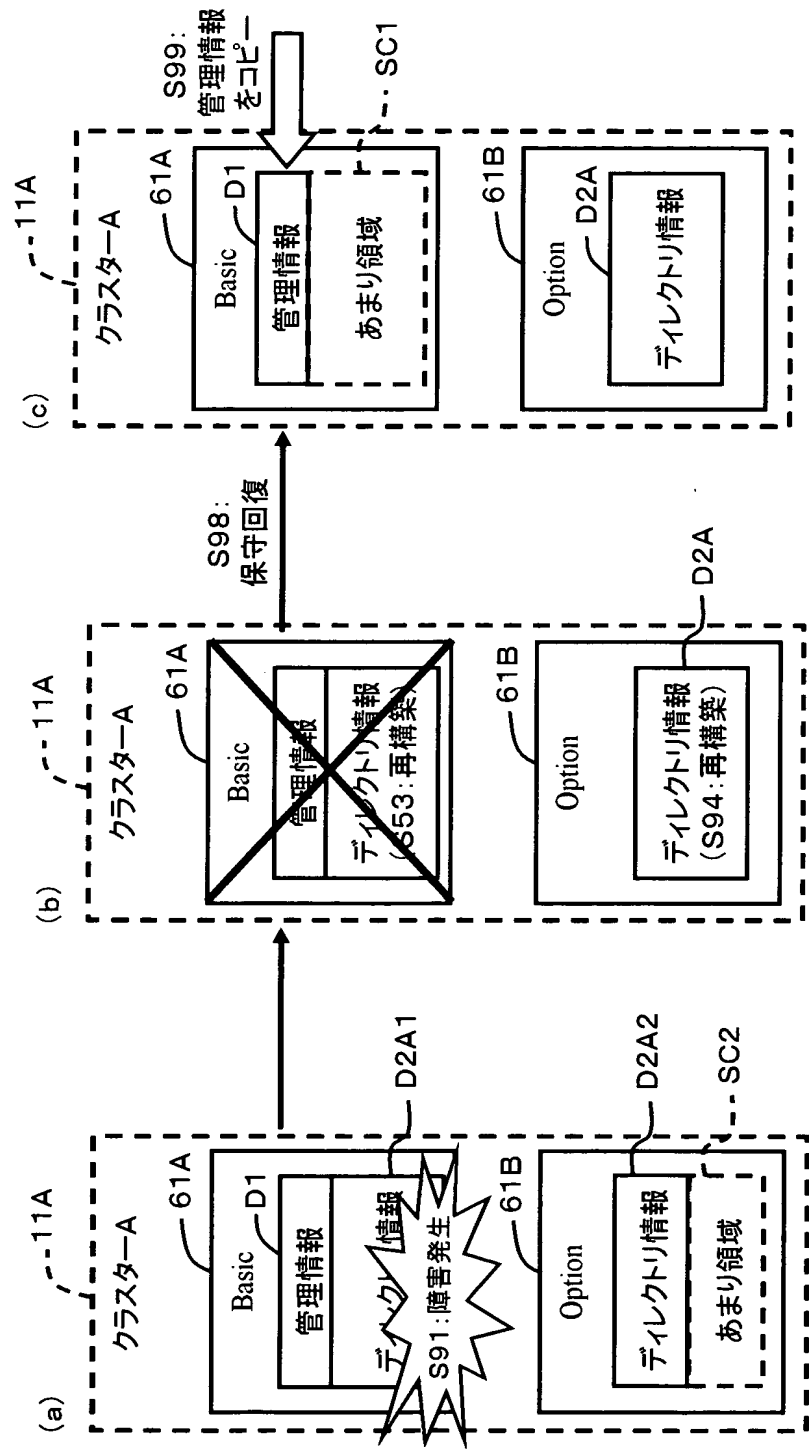
【図 18】



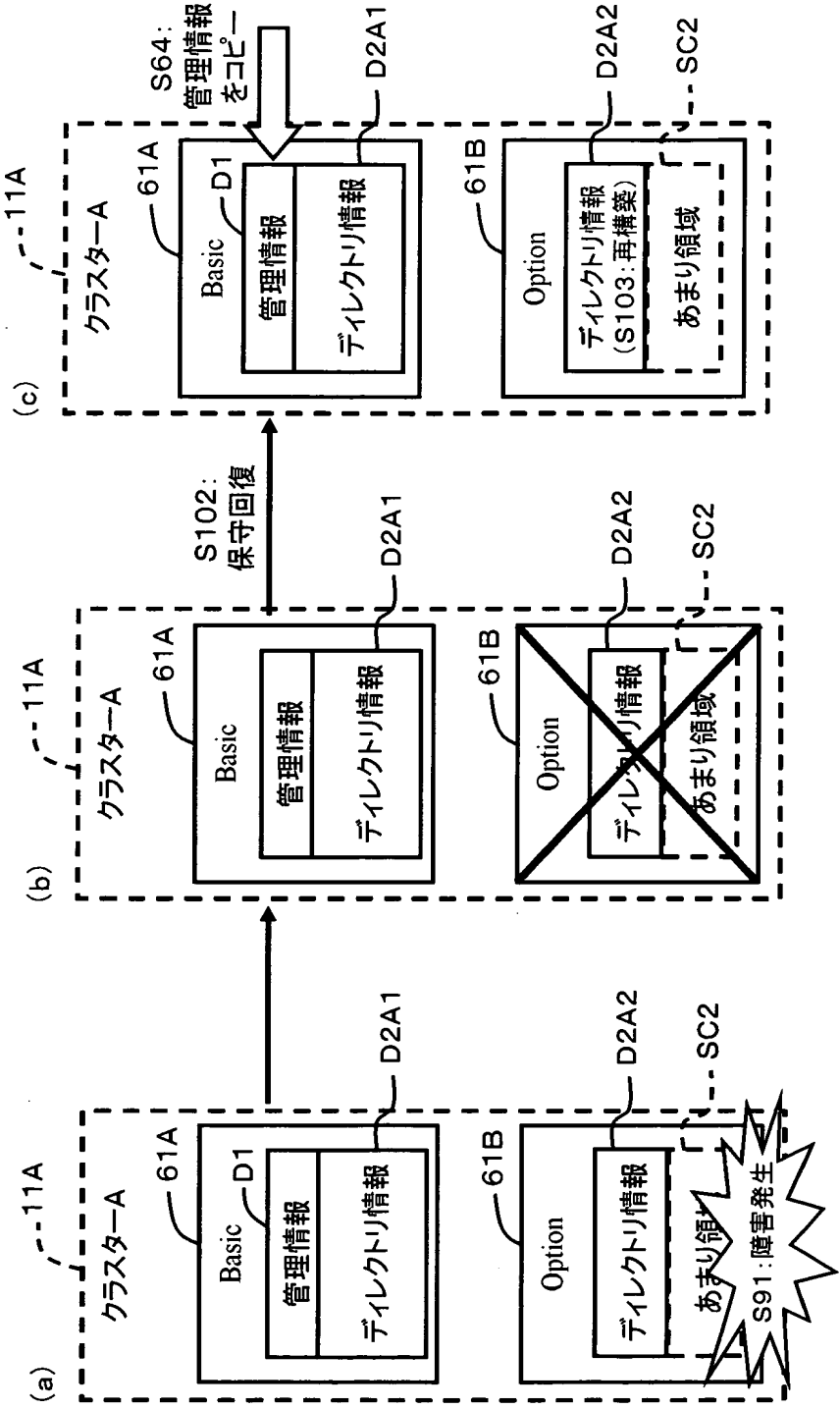
【図19】



【図 20】



【図 21】



【書類名】 要約書**【要約】**

【課題】 共有メモリに障害が発生した場合、疑似スルー動作の期間を抑制し、ディスクアレイ装置の性能低下期間を短縮する。

【解決手段】 制御情報を2重化が要求される管理情報D1と1重化で済むディレクトリ情報D2Aとに分割し、それぞれ別々の共有メモリ61A, 61Bに記憶させる。ディレクトリ情報D2Aを記憶する増設メモリ部 (Option) の共有メモリ61Bに障害が発生した場合 (S11)、基本メモリ部 (Basic) の共有メモリ61Aにディレクトリ情報D2Aを再構築する。ディレクトリ情報D2Aが再構築された時点で、疑似スルー動作が解消される。拡張メモリ部のパッケージが正常品に交換された後、改めてディレクトリ情報D2Aを再構築する (S22)。基本メモリ部の共有メモリ61Aに、他方のクラスタで管理されている管理情報をコピーし (S26)、保守回復作業を完了する。

【選択図】 図9

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 4 1 6 2 3 0
受付番号	5 0 3 0 2 0 5 9 2 4 0
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 1 2 月 1 6 日

< 認定情報・付加情報 >

【提出日】 平成15年12月15日

特願 2 0 0 3 - 4 1 6 2 3 0

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所